Measuring Executive Function During Early Childhood:

The Utility of Direct Assessments, Teacher Ratings, and Group-Based Tasks

by

Sammy F. Ahmed

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2019

Doctoral Committee

    Professor Frederick J. Morrison, Co-Chair
    Professor Pamela Davis-Kean, Co-Chair
    Professor William J. Gehring
    Assistant Professor Christina Jo Weiland

DEDICATION

To mama. Thank you and I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Among the many factors contributing to children's development, executive function (EF) skills have received a considerable amount of attention in recent years, given their role in developmental outcomes, such as academic achievement and behavior problems. Yet, much of our understanding of EF development is grounded in experimental studies that rely on highly controlled, laboratory-based measures, usually administered one-on-one to individual children. While these techniques have provided insight into the development of EF, they may not account for important contextual factors that might influence the ways in which children employ EF in important learning environments such as the classroom. EF and related cognitive and self-regulatory skills often play out in group settings, with natural distractors and social interactions with peers and teachers. Increasing the ecological sensitivity of early EF measures has the potential to enhance our understanding of how EF manifests in a naturalistic classroom setting and how it relates to specific classroom behavior and academic outcomes. Therefore, the goals of present dissertation were to 1) develop and validate a new set of EF measures, specifically designed to capture young children's EF in a dynamic social context with peers and distractors 2) compare these measures to direct assessments, and to teacher ratings of children's EF, and 3) examine the degree to which performance on EF tasks across measurement

types is related to children's academic achievement during kindergarten. Results from study 1) provide psychometric evidence that a group-based paradigm is a valid approach to studying EF processes in socially demanding contexts. Findings from study 2) revealed only modest correlations between individually assessed EF components and components derived from teacher ratings. Relations among teacher-rated EF and the group-based tasks were larger in magnitude and mapped onto corresponding sub-components across both measurement types, which might suggest that the social context exerts a similar demand on the way children employ EF skills in a classroom and group setting. Results from study 3) revealed that patterns of predictions to math and reading skills across measurement contexts were similar in significance, direction and magnitude of effect. Finally, although teacher-reported EF factors were statistically significant predictors of both math and reading achievement in kindergarten, interpretations were limited due to the high levels of multicollinearity between teacher-reported EF factors. Contributions to developmental theory are discussed.

Chapter I: Introduction

Executive functions (EF) constitute a set of broad neurocognitive processes that aid in the ability to complete tasks and purposefully guide thoughts and behaviors to achieve goals (Cartwright, 2012). This complex set of skills involves processing and manipulating information, inhibiting automatic and inappropriate responses to environmental stimuli, and directing attention to appropriate tasks (Morrison, Ponitz, & McClelland, 2010). Years of research spanning multiple academic fields have independently and consistently demonstrated the role of EF skills on an array of developmental, academic, cognitive and behavioral outcomes across various developmental stages (Ahmed, Tang, Waters, & Davis-Kean, 2018; Duncan et al., 2007; Lan, Legare, Ponitz, Li, & Morrison, 2011; Moffitt et al., 2011).

The early development of EF skills, however, has and continues to be a focal point of research given its relevance for school readiness and early academic outcomes. Specifically, complex EF skills that require the monitoring of overt, deliberate activities, are particularly useful in a learning environment where students are constantly expected to pay attention, follow rules and concentrate on various cognitive and behavioral tasks (Anderson, 2002; Blair, 2002; Blair & Razza, 2007; Diamond & Lee, 2011; Samuels, Tournaki, Blackman & Zilinski, 2016). This is especially true during the transition into formal schooling, a time when children begin to encounter growing demands associated with structured

activities and formal learning, and are asked to exhibit more self-control (Pianta & Rimm- Kaufman, 2006). For young students, this means having to wait their turn to engage in activities, raise their hand before speaking, and resist becoming distracted by peers (Rimm- Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Children's successful academic development, therefore, largely relies on their ability to control their reactions and task-related behaviors (Morrison, Ponitz, & McClelland, 2010).

Despite its demonstrated importance, much of our understanding of EF development is grounded in experimental studies that rely on highly controlled, laboratory-based measures conducted with individual children. While the use of these methods has provided tremendous insight into the development of EF, laboratory-based tasks do not account for contextual factors that might influence the ways in which children employ EF in important environments such as the classroom. EF and related cognitive and self-regulatory skills often play out in group settings, with natural distractors and social interactions with peers and teachers. Increasing the ecological sensitivity of early EF measures has the potential to enhance our understanding of how EF manifests in a naturalistic classroom setting and how it relates to specific classroom behavior and academic outcomes.

Among the several available options for measuring EF during early childhood, relatively few measures have been developed with the intention of

capturing young children's EF in a dynamic, and naturalistic context with peers, distractors, and competing social options (McCoy, 2019). Newer observational approaches and existing rater reports offer a way to assess children's EF skills in context, but are limited in the information they can provide about how children employ EF skills in socially demanding settings. Although classroom-based observational measures have the power to capture children in their natural environment, their sensitivity to specific phenomena of interest can be eroded by lack of experimental control inherent in the real-world. In addition, because researchers can only passively observe what is occurring in the classroom setting, it is often not possible to elicit specific behaviors of interest or to differentially study separate EF components (Ahmed, Grammer, & Morrison, *under review;* McCoy, 2019).

In contrast to the time-intensive nature of classroom observations, teacher report measures provide researchers with an efficient option for characterizing children's skills from the perspective of individuals with extensive experience with children. Although these types of ratings have been shown to be strong predictors of academic achievement throughout early development (e.g., McClelland, Acock, & Morrison, 2006; McClelland, Morrison, & Holmes, 2000), they are also subjective and susceptible to bias (e.g., Derks, Hudziak, & Boomsma, 2009), and evidence for consistency across teacher reports and performance-based measures is not always found in individual investigations

(e.g., Bodnar et al., 2007; Silver, 2012). Therefore, the goals of this dissertation are to 1.) introduce a new set of EF measures, specifically designed to capture young children's EF in a dynamic social context with peers and distractors. 2.) Compare these assessments to individually assessed, laboratory-based EF measures, and to teacher-reported ratings of children's EF, and 3.) examine the degree to which performance on EF tasks across these three broad measurement types is related to children's academic achievement during the early years of schooling.

Specifically, in the first study of this dissertation, I will describe and validate newly-developed group-based EF tasks. These tasks involve engaging children to complete multistep instructions in a group setting, among same-age peers, and are designed to assess children's working memory, inhibitory control, and sustained attention. This study is motivated by filling a methodological void in the literature by putting forward a set of EF tasks that take place in schools, and capture children's EF in a social context. Compared to individually assessed laboratory EF tasks, these measures can shed light onto the contextual influences exerted on children while employing EF skills. Further, unlike observational and rater approaches, which are limited in their conceptual precision (McCoy, 2019), these tasks allow researchers to operationalize and measure specific sub-components of EF in the context of peer interactions and distractions that naturally occur in group settings.

The second study explores associations among EF components across three measurement types (direct assessments, teacher ratings, and the newly developed group-based EF tasks). Given that the study of early EF development has drawn the interest of scientists from different disciplines, each with its own measurement traditions, techniques, and theoretical perspectives (Morrison & Grammer, 2016), it has become increasingly important to understand the degree to which research from distinct disciplines (e.g., Cognitive Science, Developmental Psychology, Education, etc.) inform one another and the extent to which findings from different fields converge on our understanding of early cognitive processes. This question was addressed by examining the magnitude of association across EF measurement type (direct assessments, teacher ratings, and the newly developed group-based EF tasks) and component (inhibitory control, working memory, & sustained attention), which can provide information about whether EF components across measurement contexts are tapping into similar underlying constructs.

While there is some evidence that teacher ratings and direct EF assessments are related (Fuhs, Farran, & Nesbitt, 2015; McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007), this work has not been extended to a broader range of teacher-rated and direct EF components. It is possible that direct assessments are useful when understanding an individual child's cognitive capacity, but does not capture how those skills may be employed in a classroom

setting. Conversely, teacher ratings can provide evidence of how EF skills manifest in a socially demanding context, but can also be influenced by aspects other than a child's available cognitive process (Fuhs, Farran, & Nesbitt, 2015; Toplak, West, & Stanovich, 2013). Given that a recent review by Toplak, West, & Stanovich (2013) found only modest correlations between rater reports and performance-based EF measures, it is important to understand the degree to which these different assessments capture similar cognitive processes.

Further, these findings can shed light on the ways in which testing context influences the way children employ EF skills, as it remains unclear whether EF performance in a laboratory setting is related to EF manifestations in a group or classroom setting (e.g., Bodnar et al., 2007; Silver, 2012). Finally, this set of analyses can serve as an additional test of validity of the group-based EF tasks developed in the first study. Examining the magnitude of associations between group-based EF tasks and direct assessments/teacher ratings can reveal whether these tasks more closely resemble what teachers observe in the classroom, or how children employ EF in isolation, or a combination of both.

The final study examined the degree to which performance on EF tasks across measurement types (direct assessments, teacher ratings, and the newly developed group-based EF tasks) is related to children's gains in academic achievement during the early years of schooling. While there exists a sizable literature that suggests both direct assessments and teacher ratings are positively

associated with children's early math and reading achievement (e.g., Lan et. al. 2011; Fuhs, Farran, & Nesbitt, 2015), several unanswered questions remain.

First, few studies have examined the unique contribution of all three separate EF components across measurement types to children's early math and reading skills. Given that a growing number of recent studies have demonstrated the unique and relative contribution of EF sub-components on academic outcomes using direct assessments (Ahmed et al., 2018; Morgan, Farkas, Hillemeier, Pun, & Maczuga, 2018; Nguyen & Duncan, 2019), less is known about the relative contribution of sub-components derived from teacher-rated EF assessments. This approach will allow us to compare the patterns of predictions from core EF components to early math and reading achievement across two widely used EF assessment types, and the newly developed group-based tasks. Although working memory has emerged as the strongest predictor of children's early achievement in recent investigations (e.g., Ahmed et al., 2018; Lan et al., 2011; Morgan et al., 2018; Nguyen & Duncan, 2019; Purpura, Ganley, & Schmitt, 2017), these studies have relied on direct EF assessments. It is possible that a different pattern of associations would emerge for teacher-rated or group-based EF components.

Second, it is important to understand whether relations from EF components and achievement are robust to the inclusion of important covariates. Given the importance of individual, demographic and school-level influences during early childhood, and their impact on cognitive and academic outcomes

across development (e.g., Blair, 2010; Burrage et al., 2008; Davis-Kean & Sexton, 2009; Dearing, & Tang, 2010; Hackman, Gallop, Evans, & Farah, 2015; Sarsour et al., 2011; Skibbe, Connor, Morrison, & Jewkes, 2012; Weiland, and Yoshikawa, 2013), their inclusion in examinations of EF and achievement is necessary. This will allow us to understand the degree of predictability of early EF and achievement measures while holding constant relevant covariates that can influence the interpretation of associations.

Chapter II: Background

The Development of Executive Function

Although definitions vary, executive functions mainly refer to an individual's ability to complete tasks and purposefully guide their thoughts and behaviors to achieve goals (Cartwright, 2012). These abilities emerge at early infancy and continue to develop well into early adulthood and include working memory, sustained attention, and response inhibition (Blair, 2002; Zelazo, & Carlson, 2012). Working memory is the ability to simultaneously store, manipulate and recall multiple facets of information (Baddeley, 2003; Zelazo, Carlson, & Kesek, 2008). Response inhibition refers to an individual's capacity to deliberately suppress pre-potent, automatic and impulsive responses to stimuli, and to engage in directed, deliberate and goal-oriented behavior (Friedman & Miyake, 2004). Sustained attention is the ability to focus and shift one's attention to given aspects of a task in the presence of external, and potentially distracting stimuli (Rothbart & Posner, 2005).

Work on the specific timing of EF development has been derived from various behavioral and experimental paradigms with children, and has been approached from psychological and cognitive development perspectives. Research has demonstrated three major cycles of development when there appears to be accelerated periods of growth in EF skills (18 months to 5 years; 5 to 10 years; and 10 to 14 years), which parallel models of neural and cortical

development (Durston, Davidson, Tottenham, Galvan, Spicer, Fossella, & Casey, 2006; Eslinger, Blair, Wang, Lipovsky, Realmuto, Baker, Thorne, Gamson, Zimmerman, Yang, & Rohrer, 2009; Fair, Cohen, Power, Dosenbach, Church, Miezin, Schlaggar, & Petersen, 2009).

Children's attention control begins to mature at age one and gradually develops with greater speed and accuracy by age six (Fish & Kloos, 2016; Mezzacappa, 2004). By three years of age, children display varying levels of working memory, response inhibition and the ability to shift and sustain attention (Hughes, 1998), which indicates that they can control their own behavior and attend to and remember information to complete tasks. Inhibitory control improves drastically between ages 3 and 6 (Montgomery & Koeltzow, 2010), while attentional shifting and working memory begin to markedly improve between the ages of 4 and 5, and domain general cognitive control improving between 7 and 9 years of age and continues to develop into adulthood (Jacques & Zelazo, 2001; Smidts, Jacobs, & Anderson, 2004). Like the cortical patterns of brain development, research on cognitive EF skills have suggested that that EF undergoes significant reorganization throughout development (Hughes et al. 2010; Miyake et al, 2000; Wiebe et al. 2011; Willoughby, Blair et al. 2010; Zelazo, Blair, & Willoughby, 2016).

The *differentiation hypothesis* (Garrett, 1946) proposes that the functional organization of cognitive abilities develop from a unitary and general ability

during early childhood to a more specialized, specific and differentiated set of skills by adolescence and early adulthood (e.g., Shing et al., 2010; Werner, 1957), and that these skills follow a hierarchical and integrative process throughout development. Recent studies have supported this framework by consistently yielding a unitary EF factor structure among young children up to the ages of 8 (e.g., Brydges, Reid, Fox, & Anderson, 2012; Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012; Wiebe et al., 2008), with differentiation occurring during late childhood. For example, Lehto, Juujarvi, Kooistra, and Pulkkinen (2003) found that a three-factor model with set shifting, working memory, and inhibition emerge in a sample of 8- to 13-year-olds. These, and other seminal studies have given rise to the idea that EF is better understood as a global set of skills in early childhood, that differentiate into three separate, but related constructs during the transition into late childhood and adolescence.

From a behavioral perspective, the early cycles of EF development have been a focal point of research given their relevance for school readiness and early academic achievement, especially during the transition into formal schooling. This is a time when children begin to encounter growing demands associated with structured activities and formal learning, and are asked to exhibit more EF and self-control (Pianta & Rimm- Kaufman, 2006). For young students, this means having to wait their turn to engage in activities, raise their hand before speaking, and resist becoming distracted by peers (Rimm- Kaufman, Curby, Grimm,

11

Nathanson, & Brock, 2009). As such, EF processes mature rapidly during this period and several recent studies report clear improvement in EF components during the first few years of formal schooling (e.g., Best, Miller, & Naglieri, 2011; Brydges, Lee, Bull, & Ho, 2013; Shing, Lindenberger, Diamond, Lee, & Davidson, 2010; Wiebe, Sheffield, & Espy, 2012; Willoughby et al. 2012).

Despite this accelerated period of early EF growth during the transition to school, the specific schooling factors that shape and contribute to the growth of executive function has been of recent interest and continues to be an active area of research in several scientific fields of inquiry. More specifically, parsing out the specific age and schooling effects on executive function has revealed both experiential and maturational contributors to the development of executive function. Researchers have utilized a natural experiment that allows the isolation of the effects of schooling on children's outcomes. The "school cut-off" design (Morrison, Smith, & Dow-Ehrensberger, 1995) leverages the fact that many school districts have a cutoff date for school enrollment, such that children born on or before a state mandated date can begin formal school, whereas children born after that date enter school the following year. Capitalizing on this arbitrary cut point, researchers can estimate the average treatment effect of schooling by matching children on important demographic variables (IQ, SES, home environment, etc.) and following them into the first years of schooling. This technique allows researchers to chart the growth of cognitive development and

academic achievement while parsing out environmental and maturational contributions (Christian, Bachnan, & Morrison, 2001).

Burrage and colleagues (2008) found both specific schooling and age related changes in working memory. Using this natural experiment, it was found that kindergarteners outperformed preschoolers of the same age on working memory tasks in both the fall and spring semesters ($\eta^2 = .142$). These differences in working memory performance in both the fall and spring semesters suggest that these skills improved because of schooling experiences. In other words, despite being the same age, the kindergarteners began the school year outperforming their same-age preschool counterparts due to their schooling experiences in preschool and continued to outperform them into the spring semester because of their differential schooling experience resulting from being in kindergarten (Burrage, Ponitz, McCready, Shah, Sims, Jewkes, & Morrison, 2008). Complementing these results, findings from recent investigations using the regression discontinuity design have also demonstrated the importance of early school experiences for EF skills including inhibitory control, cognitive flexibility, and working memory (Weiland, Eidelman, & Yoshikawa, 2011; Skibbe et al. 2010; Weiland & Yoshikawa, 2013).

Evidence from intervention studies have also revealed specific program effects in the growth of EF. In a recent experiment, researchers evaluated 'Tools of the Mind', which is a program intended to enhance children's behavioral

regulation through several interactive and structured classroom activities. Children in the Tools of the Mind classrooms for 1 or 2 years scored significantly higher (84 % vs. 65 %) on EF and self-regulation assessments than children with no program experience (Diamond, Bamett, Thomas, & Munro, 2007). Program Alternative Thinking Strategies (PATHS) is a curriculum add-on that trains teachers to foster children's inhibition and self-control by teaching children to recognize and manage their emotions, as well as helping them build interpersonal problem solving techniques. In their assessment, the authors found that children who received this curriculum, showed larger gains in inhibitory control ($b = .53$ $SE = .19$) than children who received the usual curriculum (Riggs, Greenberg, Kusche, & Pentz, 2006). Finally, in 2011, Raver and colleagues tested the effectiveness of the Chicago School Readiness Program (CSRP) in low-income, Head Start funded classrooms and revealed effective gains in executive function and effortful control over the course of a school year ($b = .28$ $SE = .13$). CSRP is an interventional program that helps teachers utilize classroom management strategies that help children regulate their behavior.

In addition to school cut-off and intervention approaches, observational studies have demonstrated that structured classrooms with teachers who provide organizational information about classroom rules, procedures, and activities can lead to increases in EF (Pressley, Rankin, & Yokoi, 1996). This is especially important during the school transition period, when children are learning how to

regulate themselves in a social context (Wentzel, 1991). For example, Cameron, Connor, and Morrison (2005) found that classrooms with high levels of teacher organization in the fall of the first grade fostered greater levels of EF in the fall. The findings from this study also revealed a gender interaction, such that boys benefitted more from structured classrooms with organized teachers than did girls. Together, these studies support the notion that executive function skills are, to some extent, shaped by school experiences - and that maturational growth, as a function of age, may not be the only factor in the observed acceleration of executive function growth during this developmental stage.

Importance of Executive Function for Children's Academic Achievement

Associations between EF skills and gains in academic achievement have been extensively studied in various developmental stages and across multiple domains. Years of research has demonstrated substantial relations among young children's EF skills and emerging and persistent mathematics and literacy achievement (e.g., Blair & Razza, 2007; Jacob & Parkinson, 2015; Lan et al., 2011; McClelland et al., 2007; Gathercole & Pickering, 2000). Complex EF skills that require the monitoring of overt, deliberate activities, are particularly useful in a learning environment where students are constantly expected to pay attention, follow rules and concentrate on various cognitive and behavioral tasks (Anderson, 2002; Blair, 2002; Blair & Razza, 2007; Diamond & Lee, 2011; Samuels, Tournaki, Blackman & Zilinski, 2016).

Longitudinal relations between EF and math and reading achievement

hold true for global measures of EF (e.g., Welsh, Nix, Blair, Bierman, & Nelson,

2010; Adams, Bourke, & Willis, 1999; Gathercole & Pickering, 2000), separate

components of EF (e.g., Lan et al., 2011; Morgan et al., 2019; Nguyen & Duncan,

2019), and persist across development (e.g., Ahmed et al., 2018; Blums, Belsky,

Grimm, & Chen, 2017; Bull, Espy, & Wiebe, 2008; George & Greenfield, 2005;

Hitch, Towse, & Hutton, 2001). There is a growing body of research, however,

that suggests children's EF are particularly important for supporting emerging

math skills. Many math related activities and tasks require children to actively

process and manipulate information, maintain and shift their attention to aspects

of a task, and apply reasoning based skills, all of which may require children to

activate and employ their working memory, attend to and shift their attention, and

inhibit automatic and inappropriate responses (Blair et. al., 2015; Lonigan, Allen,

& Phillips, 2017).

In 2016, Purpura Schmitt, and Ganly found associations between

inhibitory control and specific numerical abilities, such as counting, cardinality,

subitizing, and set comparison, and in a recent meta-analysis, Allan, et. al. (2014)

found a moderate association ($r = .34$) between children's inhibitory control and

performance on standardized tests of math achievement. Further, a study of

attention control showed that kindergarteners with better attention scores

outperformed students with poorer attention skills on standardized measures of

math achievement (Howse, Lange, Farran, & Boyles, 2003), and in 2015, Blair and colleagues reported that attentional control predicted growth in scores on the Applied Problems subtest of the Woodcock-Johnson Test of Achievement from preschool to second grade, over and above demographic and early achievement covariates.

Several recent studies also highlight the unique role of working memory for children's math achievement. For example, Friso-Van Den Bos, Van Der Ven, Kroesbergen, & Van Luit (2013) found moderate effect sizes between working memory, and simple and advanced arithmetic, word problems, counting, understanding of numerical concepts, and basic geometric concepts in young children. Further, in two recent longitudinal studies using nationally representative samples, kindergarten working memory emerged as the strongest EF predictor of children's math skills in third grade over and above important demographic and home-level covariates (Morgan et al., 2019; Nguyen & Duncan, 2019). Similar longitudinal trends extend to math achievement during high school. Specifically, working memory measured at 54 months was found to be the only EF component to predict math skills at age 15 in two separate investigations in large national birth cohorts (Ahmed et al., 2018; Watts, Duncan, Siegler, & Davis-Kean, 2016).

EF Measurement Traditions

Historically, researchers across academic disciplines have largely relied on three main modes of data collection when studying EF in children: direct cognitive assessments administered one-on-one to individual children, naturalistic classroom observations, and indirect teacher or parent reports. Each of these approaches offers both benefits and limitations to the study of EF.

Direct Cognitive and Behavioral Assessments

Most commonly used are direct child assessments, often adapted from standardized tasks developed by cognitive psychologists and neuroscientists. For example, the National Institutes of Health (NIH) Toolbox for the Assessments of Neurological and Behavioral Functions are a set of brief, tablet-based EF measures that are often used in clinical trials and longitudinal studies, and are suitable for children as young as 2.5 years of age. The EF measures in the NIH toolbox (Flanker task, Dimensional Change Card Sort task, and Toolbox List Sorting Working Memory Test) and other commonly used standardized EF tasks (e.g., Diamond & Taylor, 1996; Espy, 1997; Willoughby & Blair, 2011; Zelazo & Reznick, 1991) are meant to reflect the neural substrates and processes underlying specific EF components, and capture differences in children's reaction time and accuracy rates in response to stimuli. Despite the control and precision afforded by these measures, it is not clear to what extent they reflect or predict behavior in the classroom (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007).

In light of these limitations, researchers have recently begun to assess EF in school settings using direct assessments (Blair & Razza, 2007; Grammer, Carasco, Gehring, & Morrison, 2014; Ponitz, McClelland, Jewkes, Connor, Farris, & Morrison, 2008; Weiland, Barata, & Yoshikawa, 2014). However, direct behavioral measures in schools largely resemble laboratory-based tasks; they are usually administered to single students outside of the classroom, by highly trained research staff (Obradović, Sulik, Finch, & Tirado-Strayer, 2017). Although these techniques are child-friendly and are conducted in schools, they do not capture the dynamic social processes that exist in a naturalistic classroom environment. EF manifestations in a classroom occur in the context of peers and teachers, and often include distractors, challenges and competing social and behavioral options. Further, early classroom settings tend to exert demands on children to initiate behaviors, stay on-task, and monitor performance, all of which is difficult to mimic in controlled testing conditions (Silver, 2012). Ultimately, although direct behavioral assessments take place in schools, they are not designed to capture dynamic social processes, and offer limited insight into how children behave in the classroom.

Observational Approaches

Conversely, new observational approaches can offer a way to study children's EF skills as they manifest in a dynamic and naturalistic setting, such as a classroom environment. Two salient approaches to observing children's EF in

classrooms is behavioral coding of video-recorded activity during classroom

instruction, and real-time, independent assessors of children's behavior during an

activity, task or instructional practice. These newer approaches have borrowed

methods and tools from educational scientists interested in characterizing and

evaluating facets of instructional practice, teacher behavior, and learning curricula

(e.g., Individualizing Student Instruction System Connor et al., 2009; Observed

Engagement in Learning Scale; Rimm-Kaufman, 2005; Child Observations of

Preschool; Farran & Son-Yarbrough, 2001).

For example, Moffitt and Morrison (in press) developed an observational

EF measure by coding children's off-task behavior during classroom instructional

periods. Using the Individualized Student Instruction (ISI) Coding System

(Connor et al., 2009), they quantified subtypes of off-task behaviors, meant to

capture how children employ EF and self-regulation during classroom activities.

Another example of a classroom-based EF measure is the Regulation-Related

Skills Measure (RRSM; McCoy, Jones, Hemenway, Koepp, & Wilder-Smith,

2017a). This live observational approach uses Likert-scale items to rate

preschoolers' and kindergarteners' EF skills during student and teacher-led

activities and while children transition across activities in everyday classroom

environments.

While classroom-based observational measures have the power to capture

children in their natural environment, their sensitivity to the specific phenomena

of interest can be eroded by lack of experimental control inherent in the real-world. In addition, because researchers can only passively observe what is happening in the classroom setting, it is often not possible to elicit specific behaviors of interest or to differentially study separate EF components. Further, despite having high levels of ecological validity, observational approaches to studying children's EF are relatively new, less commonly used, and their psychometric properties have not been well studied or evaluated. Additionally, from a practical standpoint, these techniques take longer than direct assessments to administer, and require a substantial off-site coding investment (McCoy, 2019).

Rater Reports

EF rating scales include asking teachers, caregivers, and/or parents to rate how children employ EF related skills in typical environments, such as classrooms or home settings. The various rating scales in the literature have been shown to be effective in capturing children's behavioral EF skills from the perspective of adults with extensive experience with the child. Recent studies show that adult reports are a reliable and valid alternative to direct EF assessments in children (e.g., Gioia et al. 1996; Rothbart, Ahadi, Hershey, & Fisher, 2001; Sherman and Brooks 2010; Sulik et al. 2010).

There are various advantages to using adult ratings to assess children's EF skills. First, compared to traditional, individual assessments and observational approaches, rating scales provide researchers with an efficient, practical, and low-

cost option for assessing children's EF skills (McCoy, 2019). This is particularly

useful for early childhood researchers conducting school-based studies. Rating

scales significantly reduce testing time, resources, and are simple to administer.

Second, teacher and parent report measures allow for the characterization of

children's skills from the perspective of individuals with experience with

children, which allows raters to consider how children employ EF across a wide

range of situations, generating a general profile of children's EF skills, as opposed

to snapshots of performance on laboratory-based tasks in controlled settings

(McCoy, 2019). Relatedly, rating scales are typically considered more

ecologically sensitive than individually assessed cognitive tasks. Given that EF

and related cognitive and self-regulatory skills often play out in group settings,

with social interactions with peers and teachers, rating scales can provide

information, and capture processes that are relevant to real-world behaviors and

outcomes.

Despite these benefits, these reports are also subjective and recent studies

suggest might be susceptible to bias (e.g., Derks, Hudziak, & Boomsma, 2009).

Further, raters might have difficulty teasing apart EF skills from academic, social,

temperamental, and other behavioral constructs distinct from EF (Abikoff,

Courtney, Pelham, & Koplewicz, 1993; Duckworth & Yeager, 2015; McCoy,

2019). Issues of low precision and high subjectivity of rater reports can

potentially limit their utility in studying early EF development. In fact, recent

studies have yielded weak relations between different raters' reports on the same child using the same scale (McCoy, 2019; Mitsis, McKay, Schulz, Newcorn, & Halprin, 2000; Sullivan and Riccio 2007; Youngstrom, Loeber, & Stouthamer-Loeber, 2000), and evidence for consistency across teacher reports and performance-based measures is not always found in individual investigations (e.g., Bodnar et al., 2007; Silver, 2012).

Peer Effects on Children's Executive Function

There have been several attempts to create efficient, and ecologically valid cognitive and behavioral assessments in children. For example, McCabe and Brooks-Gunn (2007) put forward a set of three situational tasks designed to capture Delay of Gratification in group settings, with peers. These tasks mimic real-world situations that require children to regulate their behavior and are administered to groups of children in their homes and/or early care settings, such as Head Start or preschool. Specifically, the *Snack Delay* situational assessment (adapted from Kochanska, Murray, Jacques, Koenig, & Vendergeest, 1996), the *Gift Wrap* task (based on Kochanska et al., 1996), and the *Walk-a-Line* assessment (adapted from Maccoby, Dowley, Hagen, & Degerman, 1965, and more recently used by Kochanska, Murray, and Coy, 1997) were designed to measure children's Delay of Gratification outside of clinical or laboratory contexts.

The *Snack Delay* task involves asking a group of children to wait for the

administrator to ring a bell or blow a whistle before eating a treat placed in front

of the participant. This task assesses children's' ability to inhibit the urge to eat

the treat before the bell rings. Similarly, in the *Gift Wrap* task, children were

instructed not to peak while an administrator noisily wrapped a gift behind their

backs. Children were told that they would receive the gift, but that the

administrator needed to wrap it first so it would be a surprise (see McCabe et al.,

2004 for details). Like the *Snack Delay* task, children's' ability to inhibit the urge

to peak while the gift is being wrapped is also thought to capture delay of

gratification. In the *Walk-a-Line* assessment situation, four parallel (6 feet long,

and 6 inches wide) cardboard lines are placed on the floor, and children were

instructed to walk on the line normal speed, then two times as slowly from one

end of the line to the other. Start and end times for each child were coded via

video recordings of each session. Children's ability to appropriately slowdown is

thought to tap both inhibitory, and gross motor control.

These situational assessments were administered to individual children, as

well as to groups of four, same-age peers. Results revealed that across all three

assessments, children scored lower on these tasks when in the context of peers,

than when assessed individually, a finding that has implications for the way

researchers study self-regulation in young children. Since much of early

childhood research is geared toward understanding how cognitive processes affect

real-world outcomes, and child trajectories, it is critical that researchers design

instruments that accurately measure and reflect real-world processes. Although children might be able to optimally self-regulate in highly controlled lab settings, free of distractors, and challenges, those data might reveal little about how children behave in typical childhood environments, such as classrooms and early care settings. Further, children's peer groups have been shown to be important for EF and academic development (e.g., Finch, Garcia, Sulick, & Obradović, 2019; Jaccard, Blanton, & Dodge, 2005; Neidlinger, 2015), however, few standardized EF measures consider the social context in which these behaviors manifest. Thus, group-based situational tasks represent a unique experimental approach to exploring preschool children's self-regulation in a naturalistic setting, and highlight the importance of the social context when assessing children's behavior. They also draw connections between how researchers study self-regulation in the lab, and how self-regulation manifests in real-world settings.

Despite the innovation of these situational tasks and the information they provide about the importance of context for studying children's behavior, they were mainly designed to capture children's ability to delay their gratification, and are quite narrow in their scope. Other key aspects of children's self-regulatory and EF abilities can be studied using this methodological framework, and can shed light onto the social influences exerted on children's cognitive processes. Given the importance of EF components (i.e., working memory, response inhibition and attention control) for children's successful development during the school

transition (e.g., Morrison, Ponitz, & McClelland, 2010), their role in academic success (Ahmed et al., 2018; Lan et al., 2011), and how the school context shapes their development (Burrage et al., 2008; Skibbe, Connor, Morrison, & Jewkes, 2012; Weiland & Yoshikawa, 2013), it is important to understand how these skills manifest in a dynamic and naturalistic social context with peers and distractions.

The present dissertation examines extent to which behavioral EF assessments in schools can be modified to capture EF in group settings, while maintaining precision and control to study specific cognitive processes. A combination of group-based, in-school assessments can achieve a balance of both approaches. This approach will allow researchers to simultaneously capture specific EF components in the context of natural peer interactions and distractions that naturally occur in classrooms while maintaining experimental control.

Chapter III: Dissertation Studies

Study 1: Development and Validation of Group-Based EF Tasks

The first study of this dissertation describes and validates the newly developed group-based EF assessments. These tasks involve engaging children to complete multistep instructions in a group setting, among same-age peers, and are designed to assess children's working memory, inhibitory control, and sustained attention. This study is motivated by providing researchers with a set of EF tasks that take place in schools, and capture children's EF in social context. Compared to individually assessed laboratory EF tasks, these measures can shed light on the contextual influences exerted on children while employing EF skills. Further, unlike observational and rater approaches, which are limited in their conceptual precision (McCoy, 2019), these tasks allow researchers to operationalize and measure specific sub-components of EF in the context of peer interactions and distractions that naturally occur in group settings.

First, we examined the factor structure to examine the content validity and dimensionality of the group measures. Since these tasks were specifically designed to isolate and measure separate EF components in young children, we expect that a three-factor EF structure solution will best fit the data. Although multiple investigations of EF dimensionality in young children have consistently produced well-fitting, single factor solutions of EF (e.g., Hughes, Ensor, Wilson, and Graham, 2010; Wiebe et al., 2008; Wiebe et al., 2011), many of these studies

have relied on laboratory-based EF tasks, using one measure per component. Since these three situational tasks are designed to capture performance on individual EF components and are comprised of multiple sub-measures, instructions, and operations per task, we hypothesized that a three-factor solution will best fit the data. Further, given that each of the group tasks provide observable information about individual EF components, and were designed with minimal overlap across component, they have the potential to be more sensitive to early differentiation than are traditionally used EF tasks.

Next, concurrent validity was examined by comparing the group assessments with direct measures of EF. We hypothesize that children's performance on the group assessments would be significantly related to direct EF assessments. We also examined the unique predictability of each group-based EF component with analogous components derived from direct EF assessments, which would indicate that group-based EF components tap into similar constructs as standardized EF measures in young children.

Third, predictive validity of each group-based EF component was evaluated by examining relations to math and reaching achievement. We expect that children's performance on group-based EF assessments would significantly relate to standardized tests of math and reading achievement. Based on prior research (Ahmed et al., 2018; Nguyen, & Duncan, 2019; Lan et al., 2011), we expect that group-based measures of working memory to emerge as the strongest

predictors of math and reading achievement, and that all group EF components would more strongly relate to early mathematics skills than reading skills (Bull and Scerif, 2001; Cameron Ponitz et al., 2009; Bull et al., 2011).

Study 2: Relations Across EF Measurement Type

The second study of this dissertation explores associations among EF components across three measurement types (direct assessments, teacher ratings, and the newly developed group-based EF tasks). Given that the study of early EF development has drawn the interest of scientists from different disciplines, each with its own measurement traditions, techniques, and theoretical perspectives (Morrison & Grammer, 2016), it has become increasingly important to understand the degree to which research from distinct disciplines (e.g., Cognitive Science, Developmental Psychology, Education, etc.) inform one another and the extent to which findings from different fields converge on our understanding of early cognitive processes. This question will be addressed by examining the magnitude of association across EF measurement type (direct assessments, teacher ratings, and the newly developed group-based EF tasks) and component (inhibitory control, working memory, & sustained attention), which can provide information about whether EF components across measurement contexts are tapping into similar underlying constructs.

Before examining relations between EF components across measurement type, the psychometric properties and factor structure of teacher-reported EF

items will be examined. Like the group-based tasks, the teacher report scales were designed to tap into separate EF components, as reported by each student's teacher, and are comprised of multiple questions per item. Thus, we expect a three-factor solution will best fit the data.

After the latent factor(s) of the teacher-reported EF items are established, we will examine the magnitude of association across EF measurement type (direct assessments, teacher ratings, and the newly developed group-based EF tasks) and component (inhibitory control, working memory, & sustained attention).

While there is some evidence that teacher ratings and direct EF assessments are related (Fuhs, Farran, & Nesbitt, 2015; McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007), this work has not been extended to a broader range of teacher-rated and direct EF components. Given that a recent review by Toplak, West, & Stanovich (2013) found only modest correlations between rater reports and performance based EF measures, it is important to understand the degree to which these different assessments capture similar cognitive processes.

Further, these findings can shed light onto the ways in which testing context influences the way children employ EF skills, as it remains unclear whether EF performance in a laboratory setting is related to EF manifestations in a group, or classroom setting (e.g., Bodnar et al., 2007; Silver, 2012). Finally, this set of analyses can serve as an additional test of validity of the group-based EF

tasks developed in study 1. Examining the magnitude of associations between group-based EF tasks and direct assessments/teacher ratings can reveal whether these tasks more closely resemble what teachers observe in the classroom, or how children employ EF in isolation, or a combination of both.

Study 3: EF and Achievement across Measurement Type

The final study of this dissertation will be to examine the degree to which performance on EF tasks across measurement types is related to children's academic achievement. We will examine relations among direct EF assessments and math and reaching achievement during kindergarten. We expect that children's performance on group-based EF assessments would significantly relate to standardized tests of math and reading achievement. Based on prior research (Ahmed et al., 2018; Nguyen, & Duncan, 2019; Lan et al., 2011), we expect that direct measures of working memory to emerge as the strongest predictors of math and reading achievement, and that all EF components would more strongly relate to early mathematics skills (Bull and Scerif, 2001; Cameron Ponitz et al., 2009; Bull et al., 2011).

Finally, we will examine relations among teacher reported EF factors and math and reaching achievement during kindergarten. Few studies have examined the unique contribution of all three separate EF components across measurement types to children's early math and reading skills. Given that a growing number of recent studies have demonstrated the unique and relative contribution of EF sub-

31

components on academic outcomes using direct assessments (Ahmed et al., 2018; Morgan, Farkas, Hillemeier, Pun, & Maczuga, 2018; Nguyen & Duncan, 2019), less is known about the relative contribution of sub-components derived from teacher-rated EF assessments. This approach will allow us to compare the patterns of predictions from core EF components to early math and reading achievement across two widely used EF assessment types, and the newly developed group-based tasks evaluated in research question 3. Although working memory has emerged as the strongest predictor of children's early achievement in recent investigations, it is possible that a different pattern of associations would emerge for teacher-rated or group-based EF components.

Chapter IV: Method

Sample

Participants were 195 kindergarteners (104 boys, 91 girls) from four local elementary schools in Southeastern Michigan. The four schools in the study were selected to include populations of students from diverse SES and racial groups. Percentage of students eligible for free or reduced-price lunch at the school level were 2%, 61%, 68.5% and 71.9%. Forty-two percent of the sample were White/Caucasian, 33 % identified as black/African American, 17 % Asian, 2 % native American/Alaskan native, and 6 % more than one race. Children were from 4.9 to 6.9 years old (Mean = 5.7, SD = 0.38).

Procedure

Prior to testing, parents of participants and teachers provided informed consent for supervised testing. Each child was assessed by trained research assistants using a battery of individually assessed EF tasks in quiet, unused or multipurpose room in schools. Standardized tests of academic achievement were administered to all participants and the Children's Behavior Questionnaire (short version; Putnam & Rothbart, 2006) was administered to teachers to obtain ratings of each sample child's attention, working memory, and inhibitory control in a classroom setting.

Measures:

Group-based Tasks

Adapted and modified from teacher-administered challenge tasks (see Lan, 2009), the group assessments are a set of tasks designed to capture executive function (EF) performance in a naturalistic group setting. All tasks were administered and video-taped by trained research assistants, and we conducted behavioral coding to capture children's EF skills from within each task. We obtained inter-rater reliability from video recordings in the laboratory.

Freeze Task.

During the Freeze task, researchers instruct a group of five same-age participants to continually march in a circle, and to immediately freeze like a statue when music is played. When the music stops playing, children are instructed to continue marching. Children are told that they can only unfreeze themselves when the researcher says "unfreeze" or when the music stops again. A separate experimenter controls the music and starts and stops it at random intervals (7 - 15 seconds in length), using a stopwatch (see testing protocol in Appendix A). One practice trial is conducted to ensure children understand the instructions. This task is repeated for five trials and is videotaped. Reliability among scores obtained by different experimenters ranged from .87-.98. Overall average reliability was 94%.

Stop-Time is the speed with which children can stop when the music ends was coded in milliseconds.

Distractibility is the degree to which children become distracted or engage in behavior that interferes with their task performance was also coded on a Likert scale from 0 (no attention, multiple interference) to 3 (full attention, no interference).

Jumping Task.

During the Jumping task, a group of five same-age participants are told to march in a circle to music, and to freeze like a statue when the music ends. In this version of the game however, participants are given one-to-three additional instructions to complete when the music has ended. For the one-step instruction, participants are instructed to jump three times after freezing like a statue. After the one-step instruction trial, the researcher provides participants with two-step instruction ("jump three times and clap twice") and three-step instruction ("jump three times, clap twice, and crouch down"). Thus, when children march they will also have to retain the instruction for what to do once the music is over. A separate experimenter controls the music and stops it at random intervals (7 - 15 seconds in length), using a stopwatch (see testing protocol in appendix A). One practice trial is conducted to ensure children understand the instructions. Reliability among scores obtained by different experimenters ranged from .82-.97. Overall reliability was 92%.

Action Recall is children's ability to correctly follow the steps will be scored on a Likert scale from 0 (no response) to 4 (perfect recall).

Action Performance is whether a child preforms more or fewer actions than instructed. Children are given a separate score 0 (perfect performance), 1 (deviation below), and 2 (deviation above).

Stop-Time is the speed with which children can stop when the music ends was coded in milliseconds.

Distractibility is the degree to which children become distracted or engage in behavior that interferes with their task performance was also coded on a Likert scale from 3 (full attention, no interference) to 0 (no attention, multiple interference).

Marching Task.

During the **Marching Tasks**, a group of five same-age participants are told stand in a circle facing the center. They are instructed to march in one direction if one song begins playing (left for song 1) and the opposite direction if another song begins to play (right for song 2). After a few seconds, the song will stop and children are directed to turn back and face the center before starting the next trial. A separate experimenter controls the music and stops it at random intervals (7 - 15 seconds in length), using a stopwatch (see testing protocol in appendix A). One practice trial is conducted to ensure children understand the instructions. Reliability among scores obtained by different experimenters ranged from .83-.97. Overall reliability was 86%.

March Performance is the degree to which children can correctly march in the appropriate direction and was scored on a Likert scale; 0 (incorrect) 1 (self-corrected) and 2 (correct).

March Recall is the degree to which children are visually or verbally cued by a group peer and was scored as a dichotomous variable, where 0 is not cued and 1 is cued

Stop-Time is the speed with which children can stop when the music ends was coded in milliseconds.

Distractibility is the degree to which children become distracted or engage in behavior that interferes with their task performance was also coded on a Likert scale from 3 (full attention, no interference) to 0 (no attention, multiple interference).

Individually Assessed Tasks*:*

Response inhibition. The **Head-to-Toes, Knees-to-Shoulders (HTKS)** task (Ponitz et al., 2008) was used to measure children's response inhibition. In this task, the child was asked to perform opposite movements of the directions the instructor gave. For example, when asked to touch their head, children are expected to touch their toes. The task became increasingly challenging across 30 trials in three blocks (touching heads and toes, touching knees and shoulders, and a mix of the two). Incorrect responses were scored with a 0, 1 for a self-corrected response, and 2 for a correct response. The maximum score was 60. The internal

consistency in the current study (Cronbach's α) is .83. Reliability among overall scores obtained by different experimenters was 100%.

Working memory. The **Backward Digit Span** subtest of the Wechsler scales (Wechsler, 1991) was administered to gauge working memory skills, as the task requires the manipulation of information stored in memory. This task was composed of two sections in which the instructor says a list of numbers and the participant was asked to recite the numbers backwards. The list of numbers increases by one item for every correct response, and the largest set was six. If the participant answered incorrectly twice in a row, the instructor moved onto the next section (Nestbitt et al., 2013). A score was assigned based upon the largest set at which the child successfully reported. This measure demonstrates acceptable test-retest reliability (r=.73; Lipsey et al., 2017).

Attention control. The **Pair Cancellation** task, which is a measure drawn from the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock & Mather, 2000), was used to test children's attention control. In this task, children were presented with a testing sheet with small pictures of dogs, balls, and cups, and asked to circle all of the ball-dog pairs in which a dog is presented after a ball. After practicing to ensure that the child understood the task, they were given three minutes to complete the rest of the page, working as quickly as they could without making mistakes. There were 69 correct pairs, and the number of correct pairs identified within three minutes was recorded. In this study, we used W scores

representing children's ability level based on the Rasch measurement model, which provided comparative scores for each child, regardless of age. Test-retest reliability for this subtest is $r = .78$ (McGrew & Woodcock, 2001).

Academic Achievement: Math and reading skills were measured in kindergarten using the Woodcock-Johnson III Tests of Achievement Applied Problems and Letter-Word Identification subtests, respectively (Woodcock, McGrew, & Mather, 2001). The Applied Problems subtest assessed early math skills using word problems, pictures and numbers. The participants were asked to listen to the item, determine the procedure to solve the problem and successfully complete the computations. The task grew increasingly difficult and the participants were given a pencil and paper after they reached a certain point in the task to help solve the problems (Woodcock & Mather, 2000; Matthews, Ponitz, & Morrison, 2009). The Letter-Word Identification subtest assessed reading skills. Children were asked to read letters and words with proper pronunciation. The task grew more difficult as the participant progressed. W scores were also used for these two subtests in this study. Reliability and internal consistency are well documented (Woodcock & Johnson, 1990).

Teacher Reported EF Questionnaire:

Eight items from the Children's Behavior Questionnaire (short version; Putnam & Rothbart, 2006) were used to assess teachers' ratings of their student's attention and inhibitory control in a classroom setting. The Children's Behavior

Questionnaire short version consists of "My student..." statements, with 5 items measuring children's attention control (e. g., "when drawing or coloring in a book, shows strong concentration"), and 3 items measuring inhibitory control (e. g., "can wait before entering into new activities if s/he is asked to"). We used a newly developed questionnaire that assess teachers' ratings of their student's working memory capacity and consist of 3 items (e. g., "loses track during complicated tasks and may eventually abandon these tasks"). All items used a five-point Likert scale from 3 (strongly disagree) to 7 (strongly agree). Items with negative valence were reverse-coded to create consistency across all items. Cronbach's alpha across all three subscales (Attention $\alpha = .86$ Inhibitory Control $\alpha = .78$; Working Memory $\alpha = .94$) indicated good internal consistency.

Chapter V: Statistical Methods

Analysis Plan

Descriptive analyses were conducted using IBM SPSS Statistics version 24. Confirmatory factor analyses and multiple regression models were carried out using Mplus version 8 (Muthén & Muthén, 2005). We compared the fit of models with different configurations of our variables of interest based on a priori hypotheses. The best-fitting model was then chosen using the appropriate fit statistics. Non-significant χ^2 values, which assess the overall fit and the discrepancy between the sample and fitted covariance matrices, were used to indicate good fit (Browne & Cudeck, 1993). The Root Mean Square Error of Approximation (RMSEA) is a parsimony-adjusted index of goodness of fit - values less than .08 indicate good model fit (Kline, 2005). We also evaluated Comparative Fit Index (CFI), which compares the fit of a target model to the fit of an independent, or null, model, and Tucker Lewis Index (TLI) or sometimes referred to Normed Fit Index (NFI), with values between 0.95 and 1.00 indicating good model fit (Kline, 2005; Hu & Bentler, 1999). Finally, we examined the Standardized Root Mean Square Residual (SRMR), which is the difference between the residuals of the sample covariance matrix and the hypothesized model - values less than .08 indicate acceptable model fit, and values less than .06 indicate good model fit (Browne & Cudeck, 1993; Kline, 2005; Hu & Bentler, 1999).

Study 1: Development and Validation of Group-Based EF Tasks

We used a three-step procedure to determine the construct validity of the group-based EF tasks. The first step was to examine whether a single-factor latent structure best fit the data, using confirmatory factor analysis. We then compared this to the fit of two and three-factor latent structure solutions (identified in Wiebe et al., 2008). The best-fitting model was then chosen using the appropriate fit statistics, and the resulting latent factors were used in all subsequent analyses.

Next, the concurrent validity of the group-based EF factor(s) was examined using multiple regression analysis. Each lab-based EF outcome was estimated hierarchically in two steps. In the first model, lab-based EF components were regressed on each group-based EF factor in separate OLS regressions. In model two, each lab-based EF component was regressed on all three group-based EF factors in a single model to examine the unique predictability of each group-based EF factor to corresponding lab-based EF component.

Predictive validity of the group-based EF factor(s) was examined using multiple regression analysis separately for math and reading achievement. Each achievement outcome was estimated hierarchically in three steps. In model 1, kindergarten math and reading achievement were regressed on each group-based EF factor in separate OLS regressions. In model 2, kindergarten math and reading achievement were regressed on all group-based EF factors in a single model to examine the unique predictability of each group-based EF factor to math and

reading achievement. In model 3, the child's age and sex were added to the models as covariates. All analyses were clustered by site to account for any potential biases due to nesting. To reduce the possibility of Type I errors due to multiple comparisons, Holm (Holm, 1979) corrections were applied to the final models.

Study 2: Relations Across EF Measurement Type

Following our approach to study 1, we used the same three-step procedure to determine the construct validity of the teacher-reported EF scales. The first step was to examine whether a single-factor latent structure best fit the data, using confirmatory factor analysis. We then compared this to the fit of two and three-factor latent structure solutions (identified in Wiebe et al., 2008). The best-fitting model was then chosen using the appropriate fit statistics, and the resulting latent factors were used in all subsequent analyses.

Pearson's bivariate correlations were used to examine relations among group-based, and teacher reported EF latent factors, and individually assessed EF indicators (HTKS, Digit Span-Backward, & Pair Cancellation). We interpreted correlations coefficients within and across each measurement type to provide information about the how well the separate EF components are differentiated, and the magnitude of associations among each measurement type, respectively.

Study 3: EF and Achievement across Measurement Type

Relations between direct EF assessments and achievement was examined

using multiple regression analysis separately for math and reading achievement. Each achievement outcome was estimated hierarchically in three steps. In model 1, kindergarten math and reading achievement were regressed on EF measures in separate OLS regressions. In model 2, kindergarten math and reading achievement were regressed on all three EF measures in a single model to examine the unique predictability of each EF measure to math and reading achievement. In model 3, the focal child's age and sex were added to the models as covariates. All analyses were clustered by site to account for any potential biases due to nesting. To reduce the possibility of Type I errors due to multiple comparisons, Holm (Holm, 1979) corrections were applied to the final models.

We followed the same three-step procedure from studies 1 and 2 to examine relations between teacher-reported EF factors(s) and performance on standardized tests of achievement. Specifically, each achievement outcome was estimated hierarchically in three steps. In model 1, kindergarten math and reading achievement were regressed on each teacher-reported EF factor(s) in separate OLS regressions. In model 2, kindergarten math and reading achievement were regressed on all teacher-reported EF factor(s) simultaneously in a single model to examine the unique predictability of each teacher-reported EF factor(s) to measures of math and reading achievement. In model 3, the focal child's age and sex were added to the models as covariates. All analyses were clustered by site to account for any potential biases due to nesting. To reduce the possibility of Type I

errors due to multiple comparisons, Holm (Holm, 1979) corrections were applied to the final models.

Power Analysis

Confirmatory Factor Analysis

The guidelines suggested by MacCallum, Browne, and Sugawara (1996) were used to calculate the required sample size for tests of covariance structural modeling (see equation 1). For research questions 1, we expect that a sample size of 137 is necessary to detect an effect, and a minimum sample size of 200 to model the structure using a probability level of .05, an anticipated effect size of r = .20, 8 observed variables, and 3 latent variables. This will provide the desired statistical power level of .80 to detect departures from goodness of fit (MacCallum, Browne, and Sugawara, 1996). For research question 4: we expect that a sample size of 161 is necessary to detect an effect, and a minimum sample size of 256 to model the structure using a probability level of .05, an anticipated effect size of r = .20, 11 observed variables, and 3 latent variables. For all tests, we will rely on a stringent .08 RMSEA cut-off to establish "good fit". Figure 3 provides estimates of required N, based on adjustment of RMSEA values.

Figure 1: *SEM sample size calculation.*
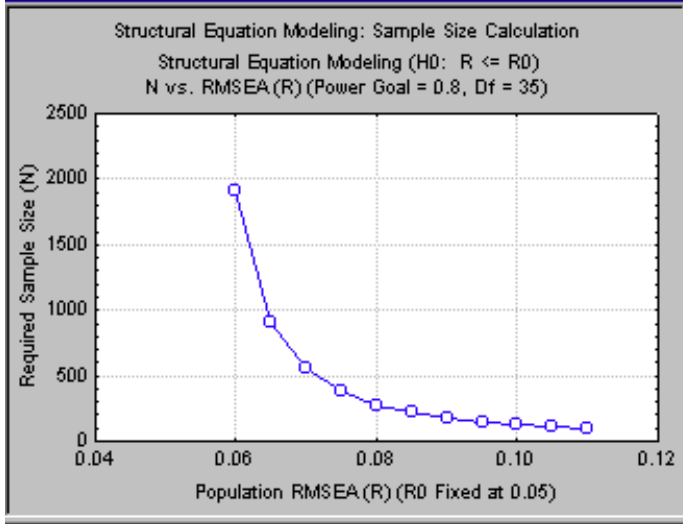
Figure 2: *Lower bound sample size for SEM.*

$$n = \max(n_1, n_2)$$

where:

$$n_1 = \left\lceil 50\left(\frac{j}{k}\right)^2 - 450\left(\frac{j}{k}\right) + 1100 \right\rceil$$

$$n_2 = \left\lceil \frac{1}{2H}\left(A\left(\frac{\pi}{6} - B + D\right) + H + \sqrt{\left(A\left(\frac{\pi}{6} - B + D\right) + H\right)^2 + 4AH\left(\frac{\pi}{6} + \sqrt{A} + 2B - C - 2D\right)}\right) \right\rceil$$

$$A = 1 - \rho^2$$

$$B = \rho \arcsin\left(\frac{\rho}{2}\right)$$

$$C = \rho \arcsin(\rho)$$

$$D = \frac{A}{\sqrt{3 - A}}$$

$$H = \left(\frac{\delta}{z_{1-\alpha/2} - z_{1-\beta}}\right)^2$$

where j is the number of observed variables, k is the number of latent

variables, ρ is the estimated Gini correlation for a bivariate normal random

46

vector, δ is the anticipated effect size, α is the Sidak-corrected Type I error rate, β is the Type II error rate, and z is a standard normal score (Cohen, 1988; Westland, 2010).

Multiple Regression Analysis

Power analysis for tests using multiple regressions analysis was carried out using G power 3.1.9.2. (Faul, Erdfelder, Lang, & Buchner, 2007). We expect small-to-medium effect sizes, ranging from .15 to .25, between EF skills and math achievement in children (Bull, & Lee, 2014; Friso-Van Den Bos, Van Der Ven, Kroesbergen, & Van Luit, 2013; Purpura, Schmitt, & Ganly, 2016). A sample size of 127 would give us 80% power to detect an effect ($r^2 = .15$), at a probability level of .05, and a total sample size of 81 to detect an effect of $r^2 = .25$.

Chapter VI: Results

Descriptive Statistics

Means, standard deviations, sample size, and range for all model indicators are presented in Table 1. To account for the impact of scaling across task type, all task scores were converted to *z*-scores before fitting our factor analytic and multiple regression models, and items and measures with negative valance were reverse-coded to aid in interpretability.

Study 1: Development and Validation of Group-Based EF Tasks

We used confirmatory factor analysis (CFA), using maximum likelihood estimation (ML) to examine the dimensionality of the group-based EF tasks. For all models, we included correlations between latent factors and controlled for both age and sex. The error variances of the variables both within and across latent factors were not correlated. The 3-factor solution of working memory, inhibitory control, and sustained attention emerged as the best fitting model ($\chi2$ (14) = 40.4, p = .000; CFI = .95; TLI = .90; SRMR = .06). As evident in Figure 1, each standardized factor loading for each latent factor was significant and exceeded a cutoff value of 0.40 (Stevens, 2001). Further, CFI, TLI, and SRMR goodness-of-fit indices of the 3-factor solution all met or exceeded the appropriate values for our determination of good model fit. The 3-factor solution did yield significant χ^2 values, however, given our sample size, classical difference testing using the χ^2 was not appropriate (Cheung and Rensvold, 2002; Chen, 2007). Both the unitary

($\chi 2$ (20) = 316.5, p = .000; CFI = .38; TLI = .14; SRMR = .14) and 2-factor ($\chi 2$ (19) = 306.0, p = .000; CFI = .40; TLI = 12; RMSEA = .14) solutions exhibited poor model fit. See table 2 for comparative fit statistics of all three CFA models.

After the best-fitting model was established for the group-based EF assessments, tests of concurrent validity were carried out in a series of step-wise regression models. Table 3 displays the results from the models predicting lab-based EF components. All group-based EF factors were significant predictors of each lab-based EF component in model 1. Next, the unique contribution of the group-based EF tasks to corresponding lab-based components were assessed by including all three group-based EF factors into a regression model simultaneously. As shown in Table 3, group-based inhibitory control continued to significantly predict both lab-based inhibition and sustained attention, whereas group-based working memory was uniquely related to direct assessments of working memory (backward digitspan), and not the other two components. Group-based sustained attention no longer predicted any of the direct EF components in model 2.

Predictive validity of the group-based EF factor was examined using multiple regression analysis separately for kindergarten math and reading achievement. Table 5 displays the results from the models predicting kindergarten math and reading achievement. All group-based EF factors were significant predictors of both math and reading achievement in kindergarten. Next, the

unique contribution of the group-based EF tasks to math and reading achievement was examined in model 2. Both inhibitory control (β = .27 SE = .10) and working memory (β = .23 SE = .08) uniquely predicted children's performance on standardized tests of math achievement in kindergarten. However, only inhibitory control (β = .33 SE = .11) continued to significantly predict reading achievement in kindergarten. In model 3, age, sex, and school fixed effects were added to the models as covariates. Both inhibitory control (β = .22 SE = .10) and working memory (β = .22 SE = .09) were robust to the inclusion of covariates, and remained statistically significant predictors of math achievement. Similarly, both inhibitory control (β = .27 SE = .12) and working memory (β = .21 SE = .10) remained statistically significant predictors of reading achievement. Group-based sustained attention no longer predicted kindergarten math (β = .10 SE = .09) or reading achievement (β = .08 SE = .10).

Study 2: Relations Across EF Measurement Type

We used confirmatory factor analysis (CFA), using maximum likelihood estimation (ML) to examine the construct validity of the teacher ratings of EF. For all models, we included correlations between latent factors and controlled for both age and sex. The error variances of the variables both within and across latent factors were not correlated. The 3-factor solution of working memory, inhibitory control, and sustained attention emerged as the best fitting model ($\chi 2$ (41) =

107.7, p = .000; CFI = .94; TLI = .92; SRMR = .046). As evident in Figure 1, each standardized factor loading for each latent factor was significant and exceeded a cutoff value of 0.40 (Stevens, 2001). Further, CFI, TLI, and SRMR goodness-of-fit indices of the 3-factor solution all met or exceeded the appropriate values for our determination of good model fit. The 3-factor solution did yield significant χ^2 values, however, given our sample size, classical difference testing using the χ^2 was not appropriate (Cheung and Rensvold, 2002; Chen, 2007). Both the unitary ($\chi2$ (45) = 186.1, p = .000; CFI = .87; TLI = .84; SRMR = .08) and 2-factor ($\chi2$ (43) = 159.1, p = .000; CFI = .89; TLI = .86; SRMR = .07) solutions exhibited poor model fit. See table 3 for comparative fit statistics of all three CFA models.

Pearson's bivariate correlations for all model indicators are presented in Table 2. With a few exceptions, there were statistically significant, positive correlations between EF components within, and across measurement type. We also examined correlations coefficients within each measurement type, which can provide some information about the how well the separate EF components are differentiated. Associations among group-based and individual EF components were moderately correlated (ranging from r = .22 to r = .47), whereas we observed large correlations among teacher reported EF components (ranging from r = .75 to r = .92). Lastly, when examining strength of associations among each measurement type, results revealed that group-based EF tasks and teacher EF

ratings are generally greater in magnitude (ranging from $r = .27$ to $r = .52$) than

those among group and individual assessments (ranging from $r = .18$ to $r = .36$).

Study 3: EF and Achievement across Measurement Type

Table 6 displays the results from the models predicting kindergarten math

and reading achievement. All individual EF measures were significant predictors

of both math and reading achievement in kindergarten. Next, the unique

contribution of the individual EF measures tasks to math and reading achievement

was examined in model 2. All three individual EF measures uniquely predicted

children's performance on standardized tests of math achievement in kindergarten.

However, only working memory and sustained attention continued to significantly

predict reading achievement in kindergarten. In model 3, age and sex were added

to the models as covariates. Both inhibitory control ($\beta = .29$ SE $= .05$) and

working memory ($\beta = .37$ SE $= .05$) were robust to the inclusion of covariates,

and remained significant predictors of math achievement and only working

memory continued to significantly predict reading achievement ($\beta = .38$ SE $=$

.06). Individual EF measures of sustained attention no longer predicted

kindergarten math ($\beta = .11$ SE $= .05$) or reading achievement ($\beta = .14$ SE $= .06$).

Table 7 displays the results from the models predicting kindergarten math

and reading achievement. All teacher-reported EF factors were significant

predictors of both math and reading achievement in kindergarten. However, when

including all teacher-reported EF factor simultaneously in a single model, we

could not interpret the estimates due to the high levels of multicollinearity

between teacher-reported EF factors.

Chapter VII: Discussion

This dissertation examined the degree to which measurement approaches from distinct disciplines converge on our understanding of EF during early childhood. Three main findings emerged from this dissertation. First, psychometric evidence suggest that a group-based paradigm is a valid approach to studying early EF processes in socially demanding contexts. Second, results reveal only modest correlations between individually assessed and teacher-rated EF components. Further, relations among teacher-rated EF and the group-based tasks were larger in magnitude and mapped onto corresponding sub-components across both measurement types, which might suggest that the social context exerts a similar demand on the way children employ EF skills in a classroom and group setting. Third, results show that patterns of predictions to math and reading skills across measurement contexts were similar in significance, direction and magnitude of effect. Finally, although teacher-reported EF factors were significant predictors of both math and reading achievement in kindergarten, interpretations were limited due to the high levels of multicollinearity between teacher-reported EF factors.

Study 1: Group-Based EF Measures

The purpose of this study was to describe and validate a new set of group-based EF assessments in young children. These efforts were motivated by filling a methodological gap in the literature by providing researchers with a set of

ecologically sensitive EF measures that capture dynamic social processes, present in group settings. Although ecologically valid approaches to studying children's behavior exist using a variety of methods (e.g., time-sampled naturalistic observations, teacher and parent reports, etc.), very often however, it is at the cost of conceptual precision and specificity. Although it is important to understand how cognitive process manifest in naturalistic social contexts, such as a classroom, it is also critical that researchers maintain the ability to isolate and study specific cognitive processes. Here, we demonstrate that behavioral EF assessments in schools can be modified to capture EF processes in a group setting, while maintaining precision to study and isolate separate sub-components of EF. Specifically, we found that group-based EF assessments were uniquely related to standardized EF measures, and exhibited predictive validity to tests of math and reading achievement. We also found that patterns of relations to academic achievement were similar for individual vs. group-based assessments in direction, significance, and magnitude of effect. Overall, this study represents a first step towards developing a set of group-based EF measures that are appropriate for use with young children.

The 3-factor solution found in the group-based tasks is a significant departure from the literature, as unitary EF models have consistently been found among young children up to the age of 9 (e.g., Brydges, Reid, Fox, & Anderson, 2012; Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012; Wiebe et al., 2008).

These and other seminal studies have given rise to the idea that EF is better understood as an undifferentiated set of skills in early childhood, that differentiate into three separate, but related constructs during the transition into late childhood, and suggest that EF undergoes significant reorganization throughout development (Hughes et al. 2010; Miyake et al, 2000; Wiebe et al. 2011; Willoughby, Blair et al. 2010; Zelazo, Blair, & Willoughby, 2016).

However, it is still unclear the degree to which this pattern of development is a function of the types of tasks used during early childhood. It is possible that these findings have been confounded by measurement impurity, and in part, been the result of overlap in the cognitive demands of performance-based EF tasks. The distinction between specific EF components might be blurred due to an overlap in working memory and inhibition task demands in early childhood (Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012; Zelazo, Blair, & Willoughby, 2016). This "measurement impurity" problem has been cited as a significant challenge to EF measurement (Miyake, Emerson, and Friedman 2000; Zelazo, Blair, & Willoughby, 2016), and has prompted researchers to begin to examine more diverse EF factor structures in early childhood. As a result, there has been a growing body of literature supporting 2-factor models of inhibition and working memory in performance-based EF tasks in children (e.g., Lerner & Lonigan, 2014; Wiebe 2011).

The componential factor structure found in the present study might be due

to the nature of the tasks and how they are coded. Specifically, these group-based situational tasks were designed to capture performance on individual EF components as they are observed in real-world settings as occurring together in the same activity. The group-based tasks are comprised of multiple sub-measures, instructions and operations per task, reducing the likelihood of significant overlap in cognitive demands across task, and are coded in a way that allows for the isolation and measurement of individual EF components. Further, given that these tasks are measured in the presence of peers, with social challenges and distractors, they might represent a "purer" measure of EF components in children. The traditional context of measurement (highly controlled, laboratory settings, individual assessments, etc.) might contribute to the measurement impurity problem, hindering our ability to accurately measure EF components in young children. These data provide preliminary evidence that ecologically sensitive, group-based EF tasks might address some of the issues facing EF measurement, and offer a way to precisely measure EF components during early childhood.

We also found evidence of concurrent validity of the group-based measures. Specifically, when examined together in a single model, group-based inhibitory control and working memory were significantly and uniquely associated with direct assessments of inhibitory control and working memory, respectively. This pattern of results suggests that group-based inhibitory control and working memory tap into similar constructs as standardized EF measures in

young children. It is important to note however, although group-based sustained attention was significantly associated with all three direct EF assessments, it did not uniquely predict the standardized, direct measure of sustained attention in our final models. However, given that all three direct EF assessments used in this study require children to employ attentional processes, the variance in the group-based sustained attention factor is likely shared across all three direct EF outcomes in our models. This result might reflect the "task impurity" problem cited in the literature (Best et al., 2009; Hughes and Graham, 2002; Miller, Giesbrecht, Muller, McInerney, & Kerns, 2012; Landis and Koch, 1977; Zelazo, Blair, & Willoughby, 2016) and represents the difficulty of developing measures that capture sustained attention processes in isolation, especially in young children. Similarly, tests of predictive validity revealed that group-based inhibitory control and working memory uniquely predict children's performance on standardized tests of math achievement, however, group-based sustained attention no longer predicted kindergarten math or reading achievement in our final models. Sustained attention assessed individually or in a group context does not uniquely predict academic skills in young children, and suggests that the unique effects of EF components on academic skills are similar using distinct measures.

Study 2: Relations Across EF Measurement Type

The purpose of the second study was to explore associations among EF

components across three measurement types (direct assessments, teacher ratings, and the newly developed group-based EF tasks). While there is some evidence that teacher ratings and direct EF assessments are related (Fuhs, Farran, & Nesbitt, 2015; McClelland, Cameron, Connor, Farris, Jewkes, & Morrison, 2007), this work has not been extended to a broader range of teacher-rated and direct EF measures, and has not compared analogous EF components across measurement types.

In general, we observed only modest correlations between individually assessed EF components and components derived from teacher ratings, which suggests that although related, teacher ratings and direct EF assessments might be capturing slightly different constructs. Further, although the group-based EF components were moderately related with individually assessed EF components, they were more strongly related to teacher EF ratings, which might indicate that the group-based EF tasks more closely resemble what teachers observe in the classroom, as opposed to how children perform in quiet and controlled settings.

We also examined correlations coefficients within each measurement type, which provided some information about the how well the separate EF components are differentiated. Associations among group-based and individual EF components were moderately correlated, whereas we observed very large correlations among teacher reported EF components, suggesting that although teachers might be accurate raters of global EF skills, they might not be able to

differentiate between the subcomponents of EF. This was also bolstered by the results from the factor analysis of the teacher-reported EF items. Like the group-based tasks, we found three distinct teacher-reported EF factors. However, unlike in the group-based tasks, we observed very large correlations among teacher reported EF factors. Further, the model fit comparisons showed only marginal improvements in fit from unitary to more complex factor structures. Together these results suggest that teacher-ratings might not be as sensitive to early differentiation as are the group-based EF tasks.

Study 3: EF and Achievement Across Measurement Type

The final study of this dissertation examined the degree to which performance on EF tasks across all three measurement types (direct assessments, teacher ratings, and the newly developed group-based EF tasks) was related to children's academic achievement during kindergarten. Few studies have examined the unique contribution of separate EF components across measurement types to children's early math and reading skills. Given that a growing number of recent studies have demonstrated the unique and relative contribution of EF sub-components on academic outcomes using direct assessments (Ahmed et al., 2018; Morgan, Farkas, Hillemeier, Pun, & Maczuga, 2018; Nguyen & Duncan, 2019), less is known about the relative contribution of sub-components derived from teacher-rated EF assessments. This approach allowed us to compare the patterns of predictions from core EF components to early math and reading achievement

across two widely used EF assessment types, and the newly developed group-based tasks evaluated in research question 3.

Both inhibitory control and working memory assessed individually were significant predictors of math achievement and only working memory continued to significantly predict reading achievement. Like the group-based EF tasks, individual EF measures of sustained attention no longer predicted kindergarten math or reading achievement in our final models.

These patterns of results are consistent with the growing body of literature demonstrating differential predictions to academic achievement across core EF components (Ahmed, et al., 2018; Nguyen, & Duncan, 2019; Lan et al., 2011), and further highlights the relative importance of working memory for emerging math and reading skills in young children. It is of note that this pattern of result is maintained across measurement type. Sustained attention assessed individually or in a group context does not uniquely predict academic skills in young children, and suggests that the unique effects of EF components on academic skills are similar using distinct measures. More research is needed to understand whether the lack of unique association between sustained attention and academic achievement points to relevance of attention for children's emerging academic skills and represents a true relation or is an artifact of measurement impurity.

Finally, although teacher-reported EF factors were significant predictors of both math and reading achievement in kindergarten, when including all

teacher-reported EF factors simultaneously into a single model, we could not

interpret the estimates due to the high levels of multicollinearity between teacher-

reported EF factors. Taken together, these findings add to the documented

limitations of rater reports. Specifically, the multicollinearity obtained in the

present data is consistent with the notion that raters have difficulty teasing apart

specific behavioral/cognitive constructs (Abikoff, Courtney, Pelham, &

Koplewicz, 1993; Duckworth and Yeager 2015; McCoy, 2019), and appear to

generate general profiles of children's EF skills, as opposed to sensitively rating

distinct cognitive and behavioral processes.

While rater reports are thought to be more ecologically sensitive than

direct assessments, issues of low precision and high subjectivity limit their utility

in studying early cognitive development. These issues might also contribute to

lack of consistency across teacher reports and performance-based measures

present in the extant literature (e.g., Bodnar et al., 2007; Silver, 2012). More work

is needed to assess the utility of teacher ratings for understanding children's

behavioral development and adaptation to school, as there might be subtle

benefits to rating scales not examined in this study. Further, while teacher ratings

provide researchers with an efficient, practical, and low-cost option for assessing

children's EF skills, data from the present study points to their limited utility for

studying EF development. Although it is important to understand how children

employ EF skills in context, it is critical to have sensitive and precise instruments.

Limitations and Future Directions

There are limitations to the current study and its findings that are important to note. First, the current study relied on cross-sectional data, which prevents us from examining the development of these skills during the early years of schooling. Although we were particularly interested in developing this set of measures for young children, especially during the transition to elementary school, longitudinal data, spanning multiple years during early childhood, are needed to further validate these measures. Moreover, given that the group-based tasks are game-like, and have been designed for young children, understanding how they can be adapted for older children is important. A longitudinal design will allow us to test the stability of these measures across time points, assess their longitudinal predictions to academic skills, and further examine componential differentiation. As the expectations of children's behavior in the classroom changes markedly across elementary school, longitudinal data would allow to us to assess the developmental appropriateness and validity of these measures across this developmental period.

Secondly, although the results suggest that group-based EF tasks are related to teacher-rated EF, more work should go into understanding whether these tasks converge with observational measures of classroom behavior. Given the limitations of rater reports, a prudent next step would be to examine the degree to which these new tasks predict observable behavior in a classroom

setting. This would provide evidence that these tasks capture the ways in which children employ EF skills while learning in a classroom context.

Third, it is important to note that although one of the main goals of this dissertation was to model specific EF components across measurement types, there is evidence that the directly assessed inhibitory control measure used in the present study (HTKS) also places demands on children's working memory, especially during the more advanced trials (Ponitz et al., 2008). Future research should include measures that better isolate children's inhibitory control to understand its development and role in children's academic success during the transition to schooling. Similarly, the attention measures used in this dissertation do not have a shifting component. Given that the ability to shift one's attention to given aspects of a task in the presence of external, and potentially distracting stimuli is a key element to the componential model of EF (Rothbart & Posner, 2005), it is important to examine whether these patterns of results extend to attention shifting/flexibility measures (e.g., Flanker task, Dimensional Change Card Sort task, etc.).

Fourth, the teacher rating findings should be interpreted with caution. Although the present dissertation reveals the limitations of teacher-reported EF, they might only apply to the items from the short form of the Child Behavior Questionnaire, and not to other commonly used rating scales. Further, although the teacher-reported items of working memory used in this study demonstrated

high levels of internal consistency, they are original items developed for this study, and have not been formally validated using psychometric tests. Future work should examine whether other commonly used teacher rating scales show a similar pattern of results.

Finally, this study lacked the type of design needed to identify causal links between variables. Although several important covariates were entered in the final models, these data are correlational, and unmeasured variables might still account for the observed effects. Secondly, the generalizability of our findings is low, given these data are not nationally representative of the population. The convenience sampling design of the present study also limits our ability to generalize these findings to the larger population of children. It is possible that schools and teachers that agreed to be part of this study possess unique features that could have biased our findings. A larger, nationally representative sample would also allow us to carry out tests of invariance across important demographic variables, such as socioeconomic status, and school level variability across broad measurement contexts.

Conclusion

Overall, this study represents a first step towards developing a set of group-based EF measures that are appropriate for use with young children. Compared to traditional, individual assessments, these group-based tasks provide researchers a way to simultaneously collect data from a group of children, which

significantly reduces testing time, and resources. Further, the materials needed for these tasks are minimal, easy to transport, and are simple to administer - making them a practical, low-cost, and scalable method of data collection for school-based studies. There are also significant empirical advantages of using more ecologically sensitive measures to understand cognitive development, and have implications for the ways researchers conceptualize and measure early cognitive processes. The present lack of contextually sensitive EF measures can hinder our understanding of how EF manifests in a naturalistic classroom setting and how it relates to specific classroom behavior and academic outcomes.

Further, findings from this dissertation can provide information about the advantages and limitations of various EF measurement approaches. Given that the recent major efforts towards understanding intervention effects on children's EF skills (e.g., Tools of the Mind; Program Alternative Thinking Strategies, Chicago School Readiness Program, etc.) rely solely on direct assessments, studying the utility of different EF measures can be of practical importance (McCoy, 2019).

Specifically, understanding and measuring how children employ EF skills in typical classroom environments might be more aligned with the goals of early EF intervention programs (Jones, Barnes, Bailey, & Doolittle, 2017). As such, future work should endeavor to understand whether context-based EF measures are more sensitive to intervention effects (McCoy, 2019), and the implications that might have for our basic understanding of cognitive development. Similarly,

more work should be devoted to understanding whether different types of EF

assessments have any bearing on the identification of at-risk children. It is

possible that certain children might struggle (or excel) with employing EF in the

context of peers and competing social options, which is not captured using

traditional direct EF assessments. Given that much of early childhood research is

motivated to better understand how cognitive processes affect real-world

outcomes, and child trajectories, it is important that we develop measures that

take this ecology into consideration.

References

Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers'
ratings of disruptive behaviors: The influence of halo effects. *Journal of
Abnormal Child Psychology, 21*(5), 519–533.

Adams, A., Bourke, L., & Willis, C. (1999). Working memory and spoken
language comprehension in young children. *International Journal of
Psychology*, *34*(5-6), 364-373. doi:10.1080/002075999399701

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2018). Executive
function and academic achievement: Longitudinal relations from early
childhood to adolescence. *Journal of Educational Psychology.* doi:
10.1037/edu0000296

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J.
(2014). Relations between inhibitory control and the development of
academic skills in preschool and kindergarten: A meta-
analysis. *Developmental Psychology*, *50*(10), 2368–2379. https://doi-
org.proxy.lib.umich.edu/10.1037/a0037493.

Altemeier, L., Jones, J., Abbott, R. D., & Berninger, V. W. (2006). Executive
functions in becoming writing readers and reading writers: Note taking
and report writing in third and fifth graders. *Developmental
Neuropsychology*, *29*(1), 161-173. doi:10.1207/s15326942dn2901_8

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71-82. doi:10.1076/chin.8.2.71.8724

Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi-org.proxy.lib.umich.edu/10.1038/nrn1201

Beaman, R., Wheldall, K., & Kemp, C. (2006). Differential teacher attention to boys and girls in the classroom. *Educational review*, *58*(3), 339-366. doi:10.1080/00131910600748406

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. doi:10.1016/j.lindif.2011.01.007

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*(2), 111-127. doi:10.1037/0003-066X.57.2.111

Blair, C. (2010). Stress and the development of self-regulation in context. *Child Development Perspectives*, *4*(3), 181-188. doi:10.1111/j.1750-8606.2010.00145.x

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x

Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. Investigators, TFLP (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter-word knowledge in the early elementary grades. *Dev. Psychol*, *51*, 459-472.

Blums, A., Belsky, J., Grimm, K., & Chen, Z. (2017). Building links between early socioeconomic status, cognitive ability, and math and science achievement. *Journal of Cognition and Development*, *18*(1), 16-40. doi:10.1080/15248372.2016.1228652

Bodnar, L. E., Prahme, M .C., Cutting, L. E., Denckla, M. B., & Mahone, E. M. (2007). Construct validity of parent ratings of inhibitory control. *Child Neuropsychology, 13*(4), 345–362. doi:10.1080/09297040600899867

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences*, *8*(12), 539-546.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions*, *154*, 136-136.

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P., Sims, B. C., Jewkes, A. M., et al (2008). Age- and schooling-related effects on executive functions in young children: A natural experiment. *Child Neuropsychology, 14*(6), 510-524. doi:10.1080/09297040701756917

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage focus editions*, *154*, 136-136.

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. doi:10.1016/j.intell.2012.05.006

Bull, E., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36-41. doi: 10.1111/cdep.12059

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273-293. doi:10.1207/S15326942DN1903_3

Bunge, S.A., Dudukovic, N.M., Thomason, M.E., Vaidya, C.J., and Gabrieli, J.D. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence From fMRI. *Neuron, 33*(2):301-311.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*, *43*(1), 61-85.

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*(4), 299-319. doi:10.1016/j.jecp.2004.01.002

Cartwright, K. (2012). Insights from cognitive neuroscience: The importance of executive function for early reading development and education. *Early Education & Development*, *23*(1), 24–36. doi:10.1080/10409289.2011.615025

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233-255. doi:10.1207/S15328007SEM0902_5

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, *14*(3), 464-504. doi:10.1080/10705510701301834

Connor, C.M., Morrison, F.J., Fishman, B.J., Ponitz, C.C., Glasney, S., Underwood, P.S., Piasta, S.B., Crowe, E.C., Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85-99.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464-465. doi:10.1126/science.1134513

Christian, K., Bachman, H. J., & Morrison, F. J. (2001). Schooling and cognitive

 development. *Environmental effects on cognitive abilities*, 287-335.

Davis-Kean, P. E. & Sexton, H. R. (2009). Race differences in parental influences

 on child achievement: Multiple pathways to success. *Merrill-Palmer*

 Quarterly 55(3), 285-318. doi:10.1353/mpq.0.0023

Dearing, E., & Tang, S. (2010). The home learning environment and achievement

 during childhood. In S. Christenson & A. L. Reschly (Eds.), *Handbook on*

 school-family partnerships (pp.131-157). New York, NY: Taylor &

 Francis.

Derks, E. M., Hudziak, J. J., & Boomsma, D. I. (2009). Genetics of ADHD,

 hyperactivity, and attention problems. In *Handbook of behavior*

 genetics (pp. 361-378). Springer, New York, NY. doi:10.1007/978-0-387-

 76727-7_25

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program

 improves cognitive control. *Science (New York, NY)*, *318*(5855), 1387

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function

 development in children 4 to 12 years old. *Science*, *333*(6045), 959-964.

 doi:10.1126/science.1204529

Duckworth, A., & Seligman, M. E. (2006). Self-discipline gives girls the edge:

 Gender in self-discipline, grades, and achievement test scores. *Journal of*

 Educational Psychology, 98, 198–208. doi:10.1037/0022-0663.98.1.198

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251.

Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Developmental Science*, *9*(1), 1–8. https://doi-org.proxy.lib.umich.edu/10.1111/j.1467-7687.2005.00454.x

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428-1446. doi:10.1037/0012-1649.43.6.1428

Eslinger, P., Blair, C., Wang, J., Lipovsky, B., Realmuto, J., Baker, D., Thorne, S., Gamson, D., Zimmerman, E., Yang, Q., and Rohrer, L. (2009). A Developmental Functional Magnetic Resonance Imaging Study of Neural Systems Subserving Relational Reasoning in Childhood and Adolescence. *Brain and Cognition, 69*: 1-10.

Eslinger, P. J., Blair, C., Wang, J., Lipovsky, B., Realmuto, J., Baker, D., … Yang, Q. X. (2009). Developmental shifts in fMRI activations during visuospatial relational reasoning. *Brain and Cognition*, *69*(1), 1–10. https://doi-org.proxy.lib.umich.edu/10.1016/j.bandc.2008.04.010

Espinet, S. D., Anderson, J. E., & Zelazo, P. D. (2013). Reflection training

    improves executive function in preschool-age children: Behavioral and

    neural effects. *Developmental Cognitive Neuroscience*, *4*, 3–15. https://doi

    org.proxy.lib.umich.edu/10.1016/j.dcn.2012.11.009

Espy, K. A., Bull, R., Kaiser, H., Martin, J., & Banet, M. (2008). Methodological

    and conceptual issues in understanding the development of executive

    control in the preschool period. In V. Anderson, R. Jacobs, P. J. Anderson

    (Eds.), *Executive functions and the frontal lobes: A lifespan*

    perspective (pp. 105-121). Philadelphia, PA: Taylor & Francis.

Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U.F., Church, J.A., Miezin,

    F.M., Schlaggar, B.L., and Petersen, S.E. (2009) Functional Brain

    Networks Develop From a ''Local to Distributed'' Organization. *PLoS*

    Computational Biology, 5(5): e1000381. doi:

    10.1371/journal.pcbi.1000381

Farran, D.C., & Son-Yarbrough, W. (2001). Title I funded preschools as a

    developmental context for children's play and verbal behaviors. *Early*

    *Childhood Research Quarterly, 16(*2), 245-262.

    http://dx.doi.org/10.1016/S0885-2006(01)00100-4

Ferguson, D. M., Lloyd, M., & Horwood, L. J. (1991). Teacher evaluations of the

    performance of boys and girls. *New Zealand Journal of Educational*

    Studies, 26, 155–163.

Finch, J. E., Garcia, E. B., Sulik, M. J., & Obradović, J. (2019). Links between classmates' and individual students' executive functions in elementary school. *AERA Open.*

Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi-org.proxy.lib.umich.edu/10.1037/0096-3445.133.1.101

Friso-van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2014). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. doi: 10.1016/j.edurev.2013.05.003

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *Journal of Educational Psychology*, *107*(1), 207–221. https://doi-org.proxy.lib.umich.edu/10.1037/a0037366

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist 1*:372–378. [PubMed: 20280373]

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British*

Journal of Educational Psychology, *70*(2), 177-194.
doi:10.1348/000709900158047

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A
neural system for error detection and compensation. *Psychological
science*, *4*(6), 385-390.

George, J. L., & Greenfield, D. B. (2005). Examination of a structured problem-
solving flexibility task for assessing approaches to learning in young
children: Relation to teacher ratings and children's achievement. *Journal
of Applied Developmental Psychology*, *26*(1), 69-84.
doi:10.1016/j.appdev.2004.10.006

Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster,
M. V. (2010). Assessment of neurological and behavioural function: The
NIH toolbox. The Lancet Neurology, 9(2), 138-139. doi:10.1016/S1474-
4422(09)70335-7

Gioia, G. A., Andrwes, K., & Isquith, P. K. (1996). *Behavior rating inventory of
executive function-preschool version (BRIEF-P)*. Odessa: Psychological
Assessment Resources.

Grammer J. K., Carrasco M., Gehring W. J., Morrison F. J. (2014). Age-related
changes in error processing in young children: A school-based
intervention. *Developmental Cognitive Neuroscience*, *9*:93–105. doi:
10.1016/j.dcn.2014.02.001

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic

    status and executive function: Developmental trajectories and

    mediation. *Developmental Science*, *18*(5), 686-702.

    doi:10.1111/desc.12246

Hair Jr., J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010) Multivariate

    Data Analysis: A Global Perspective. 7th Edition, Pearson Education,

    Upper Saddle River.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working

    memory span? Theoretical accounts and applications for scholastic

    development. *Journal of Experimental Psychology*, *130*(2), 184-198.

    doi:10.1037/0096-3445.130.2.184

Holm, S. (1979). A simple sequentially rejective multiple test procedure.

    Scandinavian Journal of Statistics, 6, 65-70.

Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and

    self-regulation as predictors of achievement in economically

    disadvantaged young children. *Journal of Experimental Education*, *71*(2),

    151–174. https://doi-

    org.proxy.lib.umich.edu/10.1080/00220970309602061

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind

    and verbal ability. *British Journal of Developmental Psychology*, *16*(2),

    233-253. doi:10.1111/j.2044-835X.1998.tb00921.x

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive

 function across the transition to school predict externalizing and

 internalizing behaviors and self-perceived academic success at 6 years of

 age. *Journal of Experimental Child Psychology*, *108*(3), 663-676.

 doi:10.1016/j.jecp.2010.06.005

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive

 function across the transition to school: a latent variable approach.

 Developmental Neuropsychology, 35(1), 20–36.

 doi:10.1080/87565640903325691

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance

 structure analysis: Conventional criteria versus new

 alternatives. *Structural equation modeling: a multidisciplinary*

 journal, *6*(1), 1-55. doi:10.1080/10705519909540118

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive Function in

 Preschool Children: Examination Through Everyday Behavior.

 Developmental Neuropsychology, 26(1), 403-422.

 doi:10.1207/s15326942dn2601_3

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.

Jaccard, J., Blanton, H., & Dodge, T. (2005). Peer Influences on Risk Behavior: An Analysis of the Effects of a Close Friend. *Developmental Psychology*, *41*(1), 135–147. https://doi-org.proxy.lib.umich.edu/10.1037/0012-1649.41.1.135

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, *20*(3), 573-591.

Jones, S. M., Barnes, S. P., Bailey, R., & Doolittle, E. J. (2017). Promoting social and emotional competencies in elementary school. *The Future of Children, 27*, 49–72.

Kim, M. H., Grammer, J. K., Marulis, L. M., Carrasco, M., Morrison, F. J., & Gehring, W. J. (2016). Early math and reading achievement are associated with the error positivity. *Developmental Cognitive Neuroscience*, *22*18-26. doi:10.1016/j.dcn.2016.09.002

Kline, R. B. (2005), Principles and Practice of Structural Equation Modeling (2nd Edition ed.). New York: The Guilford Press.

Kochanska, G., Murray, K. T., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*(2), 263–277. doi:10.2307/1131849

Kochanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vendergeest, K. (1996). Inhibitory control in young children and its role in emerging

internalization. *Child Development, 6*(2)*7*, 490–507. doi:10.2307/1131828

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Lan, X. (2009). Bridging naturalistic and laboratory measures of self-regulation: The development and validation of challenge tasks. Unpublished doctoral dissertation, The University of Michigan.

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. doi:10.1016/j.jecp.2010.11.001

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*(6), 1933–1953 doi: 10.1111/cdev.12096

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. doi:10.1007/s10862-014-9424-3

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*, 59–80.

Lewis, M. D., Lamm, C., Segalowitz, S. J., Stieben, J., & Zelazo, P. D. (2006).

Neurophysiological correlates of emotion regulation in children and adolescents. *Journal of cognitive neuroscience*, *18*(3), 430-443.

Lonigan, C. J., Allan, D. M., & Phillips, B. M. (2017). Examining the predictive relations between two aspects of self-regulation and growth in preschool children's early literacy skills. Developmental Psychology, 53(1), 63–76. http://dx.doi.org/10.1037/dev0000247

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. https://doi-org.proxy.lib.umich.edu/10.1111/j.1467-8624.2004.00745.x

Maccoby, E. E., Dowley, E. M., Hagen, J. W., & Degerman, R. (1965). Activity level and intellectual functioning in normal preschool children. *Child Development, 36*(3)*,* 761–770. doi:10.2307/1126921

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology, 101*(3), 689-704. doi:10.1037/a0014240

McCabe, L. A., & Brooks-Gunn, J. (2007). With a little help from my friends? Self- regulation in groups of young children. *Infant Mental Health Journal, 28*, 584–605. doi.org/10.1002/imhj.20155

McCabe, L. A., Rebello-Britto, P., Hernandez, M., & Brooks-Gunn, J. (2004). Games children play: Observing young children's self-regulation across

laboratory, home, and school settings. In R. DelCarmen-Wiggins & A.

Carter Eds., *Handbook of infant, toddler, and preschool mental health*

assessment pp. 491–521. New York: Oxford University Press.

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.proxy.lib.umich.edu/10.1016/j.ecresq.2006.09.003

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.

McClelland, M. M., Cameron, C. E., Wanless, S. B., Murray, A., Saracho, O. N.,

& Spodek, B. (2007). Executive function, behavioral self-regulation, and

social-emotional competence. *Contemporary Perspectives on Social*

Learning in Early Childhood Education, *1*, 113-137.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for

early academic problems: The role of learning-related social skills. *Early*

childhood research quarterly, *15*(3), 307-329.

McCoy, D. C., Jones, S. M., Hemenway, A., Koepp, A., & Wilder- Smith, O.

(2017a, November). *An Observational Measure of Regulation-Related*

Skills in the Early Childhood Classroom Setting. In Paper presented at the

meeting of the Association for Public Policy Analysis & Management, Chicago, IL.

McGrew, K. S. (2001). *Technical manual: Woodcock-Johnson III*. Itasca. IL: Riverside.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ–R technical manual*. Allen, TX: DLM.

McGrew, K. S. (1993). The relationship between the Woodcock-Johnson Psycho-Educational Battery—Revised Gf-Gc cognitive clusters and reading achievement across the life-span. In B. A. Bracken, R. S. McCallum (Eds.), *Woodcock-Johnson Psycho-Educational Battery—Revised* (pp. 39-53). Brandon, VT: Clinical Psychology Publishing Co.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development*, *75*(5), 1373–1386. https://doi-org.proxy.lib.umich.edu/10.1111/j.1467-8624.2004.00746.x

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*(3)*,* 395–423. doi:10.1080/15248372.2011.585478

Miller, M., & Hinshaw, S. P. (2010). Does childhood executive function predict

adolescent functional outcomes in girls with ADHD? *Journal of Abnormal Child Psychology*, *38*(3), 315-326. doi:10.1007/s10802-009-9369-2

Mitsis, E. M., McKay, K. E., Schulz, K. P., Newcorn, J. H., & Halperin, J. M. (2000). Parent–teacher concordance for DSM-IV attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(3), 308–313.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, USA*, *108*, 2693–2698. doi:10.1073/pnas.1010076108

Moffitt, L., & Morrison, F.J. (*in press*). Off-Task Behavior in Kindergarten: Relations to Executive Function and Academic Achievement.

Monette, S., Bigras, M., & Guay, M.-C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, *109*(2), 158–173. https://doi-org.proxy.lib.umich.edu/10.1016/j.jecp.2011.01.008

Montgomery, D. E., & Koeltzow, T. E. (2010). A review of the day–night task: The Stroop paradigm and interference control in young children. *Developmental Review*, *30*(3), 308-330.

Moriguchi, Y., & Hiraki, K. (2009). Neural origin of cognitive shifting in young children. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 6017–6021. https://doi-org.proxy.lib.umich.edu/10.1073/pnas.0809747106

Morrison, F. J., & Grammer, J. K. (2012). Conceptual clutter and measurement mayhem: Proposals for cross disciplinary integration in conceptualizing and measuring executive function. *Executive Function in Preschool*.

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. In S. D. Calkins, M. A. Bell, S. D. Calkins, M. A. Bell (Eds.), *Child development at the intersection of emotion and cognition* (pp. 203-224). Washington, DC, US: American Psychological Association. doi:10.1037/12059-011

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental psychology*, *31*(5), 789.

Miyake, A., Emerson, M.J., and Friedman, N.P. (2000). Assessment of Executive Functions in Clinical Settings: Problems and Recommendations. *Seminars in Speech and Language, 21*: 169–183.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (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. doi:10.1006/cogp.1999.0734

Muthén, B. O. & Curran, P. J. General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological methods 2,* 371 (1997).

Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical analysis with latent variables: User's guide* (pp. 1998-2012). Los Angeles: Muthén & Muthén.

Neidlinger, N. M., (2015). Do Friends Help or Hinder? Examining Co-regulation in Kindergarteners. Unpublished doctoral dissertation, The University of Michigan.

Nesbitt, K. T., Baker-Ward, L., & Willoughby, M. T. (2013). Executive function mediates socio-economic and racial differences in early academic achievement. *Early Childhood Research Quarterly, 28*(4), 774-783. doi:10.1016/j.ecresq.2013.07.005

Nguyen, T., & Duncan, G. J. (2018). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly*. doi:10.1016/j.ecresq.2018.05.006

Nie, N. H., Bent, D. H., & Hull, C. H. (1975). SPSS: Statistical package for the social sciences.

Obradović, J., Sulik, M. J., Finch, J. E., & Tirado-Strayer, N. (2017). Assessing students' executive functions in the classroom: Validating a scalable

group-based procedure. *Journal of Applied Developmental Psychology, 55*, 4-13. doi.org/10.1016/j.appdev.2017.03.003

Pianta, R. C., & Rimm-Kaufman, S. (2006). The social ecology of the transition to school: Classrooms, families, and children. In K. McCartney & D. Phillips (Eds.), *Blackwell handbook of early childhood development.* (pp. 490–507). Malden, MA: Blackwell.

Ponitz, C. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2)*,* 141–158*.* doi:10.1016/j.ecresq.2007.01.004

Posner, M. I., & Rothbart, M. K. (2005). Influencing brain networks: Implications for education. *Trends in Cognitive Sciences*, *9*(3), 99–103. https://doi-org.proxy.lib.umich.edu/10.1016/j.tics.2005.01.007

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. Elementary School Journal, 96, 363-384.

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. doi:10.1016/j.jecp.2016.08.010

Putnam, S. P., & Rothbart, M. K. (2006). Development of Short and Very Short

Forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, *87*(1), 102–112. https://doi-org.proxy.lib.umich.edu/10.1207/s15327752jpa8701_09

Rast, P. & Hofer, S. M. Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: Simulation results based on actual longitudinal studies. *Psychological methods 19,* 133 (2014).

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: self-regulation as a mediating mechanism. *Child development*, *82*(1), 362-378.

Raudenbush, S. W. & Liu, X.-F. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological methods 6,* 387 (2001).

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science*, *7*(1), 91-102.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten

classroom. *Developmental Psychology*, *45*(4), 958–972. https://doi-org.proxy.lib.umich.edu/10.1037/a0015861

Rimm-Kaufman, S., Le Para, K.M., Downer, J.T., & Pianta, R.C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *Elementary School Journal, 105*(4), 377. https://doi.org/0.1086/429948

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*(5), 1394–1408.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040. https://doi-org.proxy.lib.umich.edu/10.1016/j.neuropsychologia.2003.12.012

Rypma, B., & Prabhakaran, V. (2009). When less is more and when more is more: The mediating roles of capacity and speed in brain–behavior efficiency. *Intelligence*, *37*(2), 207–222. https://doi-org.proxy.lib.umich.edu/10.1016/j.intell.2008.12.004

Rueda, M.R., Rothbart, M.K., McCandliss, B.D., Saccomanno, L., and Posner, M.I. (2005). Training, Maturation, and Genetic Influences on the Development of Executive Attention. *Proceedings of the National Academy of Sciences, 102*(41): 14931-14936.

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society*, *17*(1), 120-132. doi:10.1017/S1355617710001335

Samuels, W. E., Tournaki, N., Blackman, S., & Zilinski, C. (2016). Executive functioning predicts academic achievement in middle school: A four-year longitudinal study. *The Journal of Educational Research*, *109*(5), 478-490. doi:10.1080/00220671.2014.979913

Sherman, E. M., & Brooks, B. L. (2010). Behavior rating inventory of executive function–preschool version (BRIEF-P): Test review and clinical guidelines for use. *Child Neuropsychology, 16*(5), 503–519.

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, *35*(6), 679-697. doi:10.1080/87565641.2010.508546

Short, S.J., Elison, J.T., Goldman, B.D., Styner, M., Gu, H., Connelly, M., Maltie, E.., Wilson, S., Lin, W., Gerig, G., Resnick. J.S., and Gilmore, J.H. (2013). Associations Between White Matter Microstructure and Infants' Working Memory. *NeuroImage, 64*: 156-166. doi:10.1016/j.neuroimage.2012.09.021

Silver, C. H. (2014). Sources of data about children's executive functioning: review and commentary. *Child Neuropsychology*, *20*:1–13. doi: 10.1080/09297049.2012.727793.

Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2010). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly, 26*(1), 42–49. doi:10.1016/j.ecresq.2010.05.001

Smidts, D. P., Jacobs, R., & Anderson, V. (2004). The object classification task for children (OCTC): A measure of concept generation and mental flexibility in early childhood. *Developmental Neuropsychology*, *26*(1), 385-401. doi:10.1207/s15326942dn2601_2

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59*(4), 745-759. doi:10.1080/17470210500162854

Stipek, D., & Valentino, R. A. (2015). Early childhood memory and attention as predictors of academic growth trajectories. *Journal of Educational Psychology*, *107*(3), 771.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103. doi:10.1207/S15327752JPA8001_18

Sullivan, J. R., & Riccio, C. A. (2007). Diagnostic group differences in parent and

    teacher ratings on the BRIEF and Conners' scales. *Journal of Attention*

    Disorders, 11(3), 398–406.

Sulik, M. J., Huerta, S., Zerr, A. A., Eisenberg, N., Spinrad, T. L., Valiente, C., ...

    Edwards, A. (2010). The factor structure of effortful control and

    measurement invariance across ethnicity and sex in a high-risk sample.

    Journal of Psychopathology and Behavioral Assessment, 32(1), 8–22.

Tominey, S. L., & McClelland, M. M. (2011). Red light, purple light: Findings

    from a randomized trial using circle time games to improve behavioral

    self-regulation in preschool. *Early Education and Development*, *22*(3),

    489–519. https://doi-

    org.proxy.lib.umich.edu/10.1080/10409289.2011.574258

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Do performance-based

    measures and ratings of executive function assess the same

    construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143.

    https://doi-org.proxy.lib.umich.edu/10.1111/jcpp.12001

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning

    in children, and its relations with reasoning, reading, and

    arithmetic. *Intelligence*, *35*(5), 427–449. https://doi-

    org.proxy.lib.umich.edu/10.1016/j.intell.2006.09.001

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the Marshmallow

Test: A conceptual replication investigating links between early delay of

gratification and later outcomes. *Psychological Science*, 1–19,

doi:10.177/0956797618761661

Wanless, S. B., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2011). The

influence of demographic risk factors on children's behavioral regulation

in prekindergarten and kindergarten. *Early Education and Development,*

22(3), 461-488. doi:10.1080/10409289.2011.536132

Wechsler, D. (1991). Manual for the Wechsler intelligence scale for children-

(WISC-III). *San Antonio, TX: Psychological Corporation*.

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. doi:10.1037/a0016738

Weiland, C., Barata, M. C., & Yoshikawa, H. (2014). The co-occurring

development of executive function skills and receptive vocabulary in

preschool- aged children: A look at the direction of the developmental

pathways. *Infant and Child Development, 23*(1)*,* 4–21.

doi:10.1002/icd.1829

Weiland, C., Eidelman, H., & Yoshikawa, H. (2011). A Regression Discontinuity

Analysis of the Impact of. *Society for Research on Educational*

Effectiveness.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112-2130. doi: 10.1111/cdev.12099.

Werner, H. The concept of development from a comparative and organismic point of view. In: Harris, DB., editor. *The concept of development: An issue in the study of human behaviour*. Minneapolis: Jones Press, Inc; 1957. p. 125-148.

Wiebe, S. A., Epsy, K. A., & Charak, D. (2008). Using Confirmatory Factor Analysis to Understand Executive Control in Preschool Children: I. Latent Structure. *Developmental Psychology, 44*(2): 575

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development*, *83*(4), 1245-1261. doi:10.1111/j.1467-8624.2012.01765.x

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. doi:10.1016/j.jecp.2010.08.008

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: psychometric properties

and criterion validity of a new battery of tasks. *Psychological assessment*, *22*(2), 306. doi:10.1037/a0018708

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: longitudinal measurement invariance and developmental change. *Psychological assessment*, *24*(2), 418.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery – Revised*. Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Achievement*. DLM Teaching Resources.

Woodcock, R. W., & Mather, N. (2000). Woodcock Johnson psycho-educational battery-III, Riverside. *Itasca, IL*.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Examiner's manual. *Woodcock-Johnson III Tests of Achievement*.

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*1-9. doi:10.1016/j.lindif.2012.10.004

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting*

and Clinical Psychology, 68(6), 1038.

Zelazo, P.D. (2015). Executive Function: Reflection, Iterative Reprocessing,

Complexity, and the Developing Brain. *Developmental Review, 38*: 55–68.

doi: 10.1016/j.dr.2015.07.001

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive function:*

Implications for education (NCER 2017–2000). Washington, DC: U.S.

Department of Education, Institute of Education Sciences, National Center

for Education Research.

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in

childhood and adolescence: Development and plasticity. *Child*

*Development Perspectives*, *6*(4), 354-360.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive

function in childhood.

Zelazo, P.D., and Cunningham, W. (2007). Executive Function: Mechanisms

Underlying Emotion Regulation. In J. Gross (Ed.), Handbook of Emotion

Regulation (pp. 135-158). New York: Guilford.

Table 1.
Means, SD, Sample Size, and Range for All Indicators

Variable	N	M	SD	Min	Max
Group-Based EF Tasks					
Freeze Game					
Stop Time	195	4421.4	1810.3	1625	13539
Distractibility	195	13.1	2.7	5	15
Jumping Game					
Action Recall	195	15.1	6.1	0	68
Action Performance	195	4.6	5.3	0	36
Stop Time	195	2321.1	1263.3	940	11484
Distractibility	195	7.5	3.2	3	43
Marching Game					
March Recall	194	2.8	1.9	0	7
March Performance	194	8.0	4.1	0	14
Stop Time	194	6305.9	2099.8	2504	16394
Distractibility	194	15.5	4.8	6	22
Laboratory EF Tasks					
HTKS	215	29.2	17.1	0	58
Backward Digit Span	214	2.0	1.7	0	8
Pair-Cancellation	201	103.8	10.0	62	123
Teacher-Rated Items					
CBQ-1	133	5.3	1.4	3	7
CBQ-2	133	5.4	1.3	3	7
CBQ-3	133	5.6	1.2	3	7
CBQ-4	133	5.8	1.0	3	7
CBQ-5	133	5.4	1.3	3	7
CBQ-7	133	5.7	1.1	3	7
CBQ-9	133	5.3	1.4	3	7
CBQ-12	133	5.6	1.2	3	7
Original Item-1	133	5.4	1.2	3	7
Original Item-2	133	5.4	1.2	3	7
Original Item-3	132	5.5	1.2	3	7
Academic Achievement					
Mathematics	211	105.4	16.6	31	155
Reading	213	113.2	14.3	67	164

Figure 3: *Confirmatory Factor Analysis of Group-Based EF Tasks*
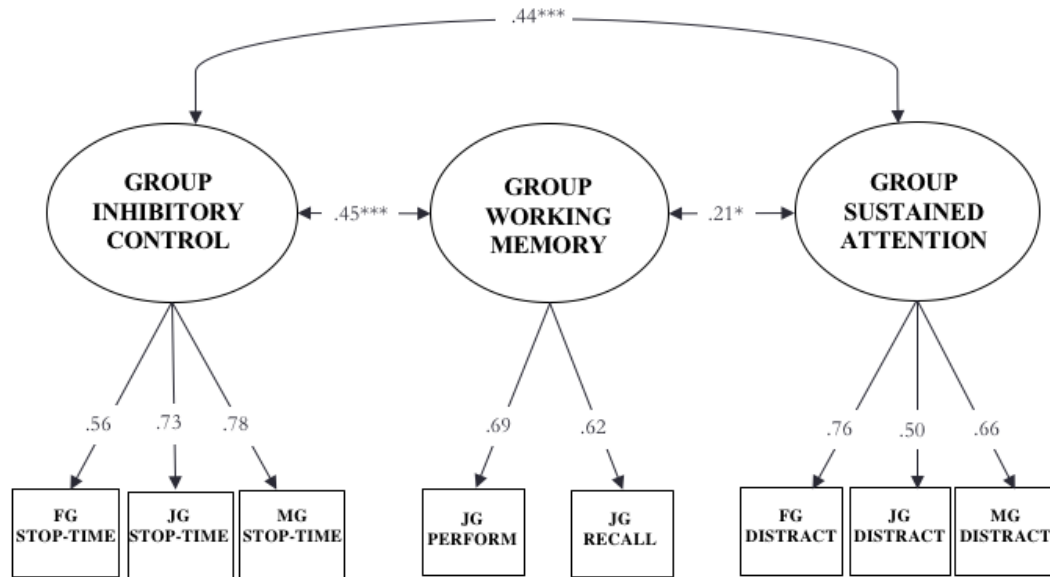
Model fit Statistics: $\chi^2(14) = 40.4$, $p = .000$; CFI = .95; TLI = .90; SRMR = .06
*$p < 05$. **$p < .01$, ***$p < .001$

Table 2: *Model Fit Indices for CFA of Group Tasks (N = 195)*

Model (Number of Factors)	*df*	χ^{2a}	χ^2/df [b]	*CFI*[c]	*SRMR*[d]	*AIC*[e]	*Model Comparison*	Δ*df*	Δχ^2
1. Unitary (1)	20	316.48*	15.82	.38	.14	4201.22	-		
2. WM & IC (2)	19	306.00*	16.11	.40	.14	4192.75	Model 1 vs. *Model 2*	1	10.48
3. WM, IC, & SA (3)	14	40.39*	2.89	.95	.06	3937.14	Model 2 vs. *Model 3*	5	265.61

Note: The preferred models are italicized; WM = Working Memory; IC = Inhibitory Control; SA = Sustained Attention
*p < .001

Table 3

Stepwise Regressions of Group EF Tasks Predicting EF Measures (N =195)

| | Kindergarten HTKS | | Kindergarten Digit Span-B | | Kindergarten Pair Cancellation | |
	Model 1[a]	Model 2[b]	Model 1[a]	Model 2[b]	Model 1[a]	Model 2[b]
Group-Based EF	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
Inhibitory Control	.33(*.07*)***	.28(*.11*)**	.38(*.07*)***	.19(*.10*)	.47(*.09*)***	.30(*.14*)*
Working Memory	.19(*.08*)**	.02(*.10*)	.37(*.08*)***	.23(*.09*)**	.29(*.10*)**	.16(*.11*)
Sustained Attention	.26(*.08*)***	.10(*.10*)	.26(*.08*)***	.11(*.09*)	.30(*.10*)***	.14(*.11*)

Note. *05. **$p < .01$, ***$p < .001$
[a]Linear regression estimates (no covariates)
[b]Multiple regression estimates (simultaneous Group-Based EF predictors)
p-values adjusted for multiple comparisons using Holm/Bonferroni correction

Table 4

Stepwise Regressions of Group EF Tasks Predicting Achievement (N =195)

Group-Based EF	Kindergarten Math[d]			Kindergarten Literacy[e]		
	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
Inhibitory Control	.45(.07)***	.27(.10)***	.22(.10)*	.46(.07)***	.33(.11)***	.27(.12)*
Working Memory	.40(.07)***	.23(.08)***	.22(.09)**	.40(.08)***	.19(.10)	.21(.10)*
Sustained Attention	.29(.08)***	.10(.10)	.10(.09)	.27(.09)***	.04(.10)	.08(.10)

Note. *p* < 05. **p* < .01, ***p* < .001

[a]Linear regression estimates (no covariates)

[b]Multiple regression estimates (all Group-Based EF predictors)

[c]Full-Model estimates with additional covariates: Age at Testing, Sex, and School Attending

[d]Kindergarten Math = The Woodcock-Johnson Applied Problems Subtest

[e]Kindergarten Literacy = The Woodcock-Johnson Letter-Word Identification subtest

p-values adjusted for multiple comparisons using Holm/Bonferroni correction

Figure 4: *Confirmatory Factor Analysis of Teacher-Reported EF*
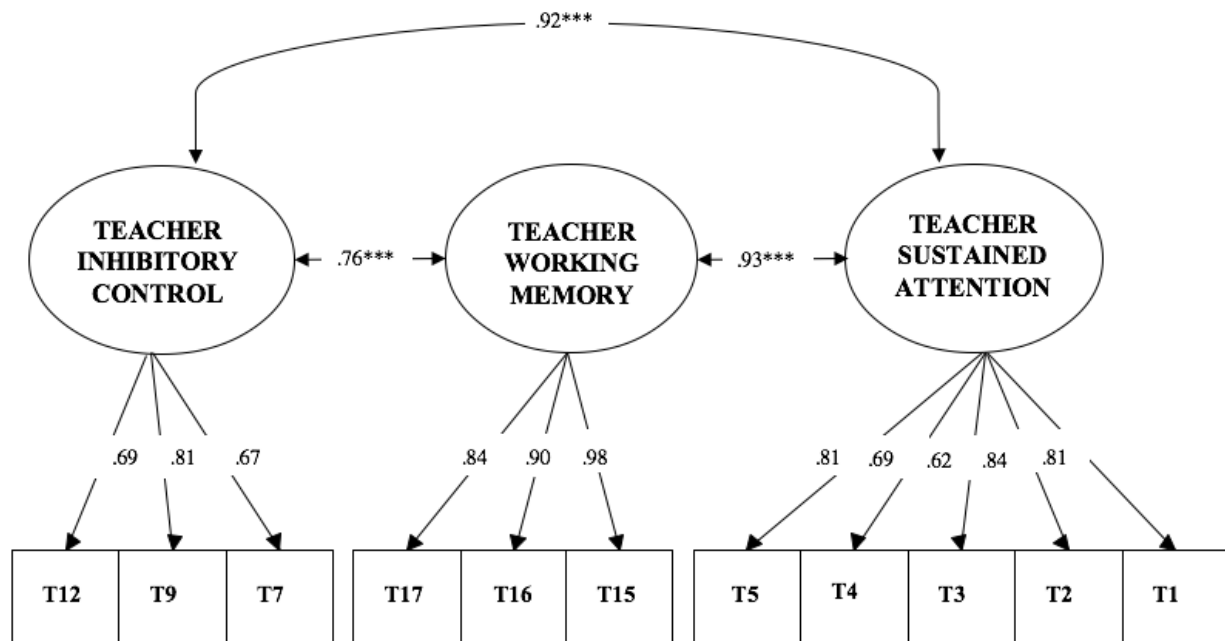
Model fit statistics: $\chi2$ (41) = 107.7, p = .000; CFI = .94; TLI = .92; SRMR = .046.
*p < .05. **p < .01. ***p < .001.

Table 5: *Model Fit Indices for CFA of Teacher Reported EF (N = 195)*

Model (Number of Factors)	df	χ^{2a}	χ^2/df [b]	CFI^c	$SRMR^d$	AIC^e	Model Comparison	Δdf	$\Delta\chi^2$
1. Unitary (1)	45	186.10*	4.14	.87	.08	3796.15	-		
2. WM & IC (2)	43	159.10*	3.70	.89	.07	3773.13	Model 1 vs. *Model 2*	2	27.10
3. WM, IC, & SA (3)	41	107.70*	2.63	.94	.05	3725.80	Model 2 vs. *Model 3*	2	51.40

Note: The preferred models are italicized; WM = Working Memory; IC = Inhibitory Control; SA = Sustained Attention
*p < .001

Table 6: *Correlations for Individual, Group, and Teacher Reported EF (N = 195)*

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. Group Inhibition	-								
2. Group Working Memory	.47***	-							
3. Group Attention	.46***	.22*	-						
4. Teacher Inhibition	.27*	.34**	.52***	-					
5. Teacher Working Memory	.41***	.42***	.38***	.75***	-				
6. Teacher Attention	.44***	.38***	.47***	.92***	.92***	-			
7. HTKS	.34***	.18	.23**	.34***	.43***	.42***	-		
8. Backward Digit Span	.35***	.34***	.25***	.04	.29***	.20*	.40***	-	
9. Pair Cancellation	.36***	.26*	.26**	.10	.38***	.35***	.26***	.28***	-

Note: * p < 05. **p < .01, ***p < .001

Table 7

Estimates of Directly Assessed EF Predicting Achievement (N =269)

	Kindergarten Math[d]			Kindergarten Literacy[e]		
	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Lab EF	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
K_HTKS	.45(*.05*)***	.29(*.05*)***	.29(*.05*)***	.25(*.06*)***	.07(*.06*)	.08(*.06*)
K_Digit Span-B	.55(*.04*)***	.39(*.05*)***	.37(*.05*)***	.49(*.04*)***	.40(*.06*)***	.38(*.06*)***
K_ Pair Cancellation	.30(*.06*)***	.13(*.05*)*	.11(*.05*)	.28(*.06*)***	.16(*.06*)**	.14(*.06*)

*Note. * p < 05, **p < .01, ***p < .001*
[a]Linear regression estimates (no covariates)
[b]Multiple regression estimates (all direct EF predictors)
[c]Full-Model estimates with additional covariates: Age at Testing, Sex, and School Attending
[d]Kindergarten Math = The Woodcock-Johnson Applied Problems Subtest
[e]Kindergarten Literacy = The Woodcock-Johnson Letter-Word Identification subtest
p-values adjusted for multiple comparisons using Holm/Bonferroni correction

Table 8

Estimates of Teacher-Reported EF Predicting Achievement (N =131)

	Kindergarten Math[d]			Kindergarten Literacy[e]		
	Model 1[a]	Model 2[b]	Model 3[c]	Model 1[a]	Model 2[b]	Model 3[c]
Teacher Reports	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
Inhibitory Control	.27(.*10*)***	-	-	.23(.*11*)*	-	-
Working Memory	.33(.*09*)***	-	-	.33(.*10*)***	-	-
Sustained Attention	.30(.*09*)***	-	-	.29(.*10*)**	-	-

Note. *p < 05, **p < .01, ***p < .001

[a]Linear regression estimates (no covariates)

[b]Multiple regression estimates (all Teacher Reported EF predictors)

[c]Full-Model estimates with additional covariates: Age at Testing, Sex, and School Attending

[d]Kindergarten Math = The Woodcock-Johnson Applied Problems Subtest

[e]Kindergarten Literacy = The Woodcock-Johnson Letter-Word Identification subtest

p-values adjusted for multiple comparisons using Holm/Bonferroni correction

Appendices

Appendix A:
Group-Based Inhibitory Control Task Coding Scheme

Stop Time
- Stop time is calculated **in millisecond**s from when the song begins playing to when the child stops and freezes

Distractibility/Self-Interference
3 –Full Attention, No Interference (body and face are directed at the experimenter, attentive expression, no disruptive actions or verbalizations)
2 –Distracted by peer (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
1 –Distracted by self (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
0 –No Attention, Multiple Instances of Interference

Trial Time
These recordings will vary by group, but will be consistent within groups
- Trial time is calculated **in milliseconds**
- For the first trial, this is calculated from when the children are instructed to begin marching to when first song segment stops playing
- For subsequent trials, this is calculated from when the song segment stops playing to when the following song segment stops playing

高

Appendix B:
Group-Based Working Memory Task Coding Scheme

Stop Time
- Stop time is calculated **in milliseconds** from when the song stops playing to when the child freezes prior to completing actions

Distractibility/Self-Interference
3 –Full Attention, No Interference (body and face are directed at the experimenter, attentive expression, no disruptive actions or verbalizations)
2 –Distracted by peer (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
1 –Distracted by self (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
0 –No Attention, Multiple Instances of Interference

Action Recall
4 –Perfect Recall (given to a child who performs the actions correctly without learning the actions from a peer)
3 –Cued Perfect (given to a child who performs the actions correctly after learning (visually, verbally) from a peer; there will be a delay in performing the actions)
2 –Incorrect (given to a child who performs the actions incorrectly without being cued by a peer)
1 –Cued Incorrect (given to a child who performs the actions incorrectly after learning (visually, verbally) from a peer; there will be a delay in performing the actions)
0 –No Response (given to a child who does not respond or produces an irrelevant or completely incorrect response)

Action Performance
*Performance will be calculated for **each** action during each trial (e.g., Trial 1: 1 action, Trial 2: 2 actions, etc.)*
Deviation Below– given to a child who performs the action(s) fewer times than were instructed (e.g., instructed to jump three times; jumps twice—given score of -1; jumps once—given score of -2, etc.)
 0 – Perfect performance (given to a child who performs the action the correct amount of times; e.g., jumps three times, claps twice and crouches down)

Deviation Above– given to a child who performs the action(s) more times than were instructed (e.g., instructed to jump three times; jumps four times—given score of 1; instructed to clap twice; claps four times—given score of 2, etc.

Marching Direction Time
-   Marching Direction time is calculated **in milliseconds** from when the song begins playing to when the child begins marching

Stop Time
-   Stop time is calculated **in milliseconds** from when the song stops playing to when the child freezes (this will be followed by "reset" motion)

Distractibility/Self-Interference
3 –Full Attention, No Interference (body and face are directed at the experimenter, attentive   expression, no disruptive actions or verbalizations)
2 –Distracted by peer (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
1 –Distracted by self (Partial Attention, Some Instances of Interference (some time with full attention, some occurrences of distraction or disruptive actions or verbalizations)
0 –No Attention, Multiple Instances of Interference

Performance (Marching Direction)
2 –Correct (given to a child who marches in the correct direction)
1 – Self- Correct (given to a child who corrects to the appropriate direction)
0 –Incorrect (given to a child who marches in the incorrect direction)

Recall (Marching Direction)
1 –Not Cued (given to a child who performs the action without learning the actions from a peer)
0 –Cued (given to a child who performs the action after learning (visually, verbally) from a peer; there will be a delay in performing the actions)

Supplies
- Speakers
- Music CD/MP3 Player
- Protocol sheet
- Labels for children
- Class list for matching children and numbers

Freeze Task:
During the Freeze Prime game, researchers instruct children to march in a circle without music. When the music starts, children must freeze into a certain pose. Children can only unfreeze themselves when the researcher says "unfreeze" or when the music stops again. A separate researcher controls the marching music and starts it at random intervals of less than 15 seconds. The task is repeated for three trials and is videotaped.
Researcher directions:
1. Get the children in a circle.
2. Say, "We are going to play a game called **Freeze**. In this game, you are going to walk in a circle (model by walking around the outside of the circle clockwise) when the music is **not** playing. When the music **starts playing**, freeze like a statue. Then when the music **stops** again, you can unfreeze and start walking around in a circle again."
3. Conduct one practice trial. If most of students understand instructions, proceed with game. If not, reiterate the rules on the game **only once**, and then proceed to game.

Researcher directions:
1. Make sure both cameras are still recording.
2. Control the music, letting children march for 7-15 seconds at a time and then starting the music.

Supplies
- Speakers
- Music CD/MP3 Player
- Protocol sheet
- Labels for children
- Class list for matching children and numbers

Jumping Task:
During the Jumping game, children also march in a circle to music, but prior to marching, researcher instruct students to "jump three times" (one-step instruction) when they hear the music stop. So, as children march, they have to monitor the music and remember the instructions. After the one-step instruction trial, the researcher gives the two-step instruction ("jump three times and clap twice") and three-step instruction ("jump three times, clap twice, and crouch down") respectively.

Researcher directions:
4. Get the children in a circle again.
5. Say, "Show me what it means to clap three times" and model it correctly.
6. Say, "Show me what it means to jump once" and model it correctly.
7. Say, "We are going to play another game. You are going to walk in a circle when the music is playing. When the music stops, **jump three times** and then stop."
8. After the children do that, say, "Good job! Now let's try it another way. This time, when the music stops, **jump three times and clap your hands twice**, then stop."
9. After the children do that, say, "Great! Now let's try one more thing. This time, when the music stops, **jump three times, clap your hands twice, and crouch down**, then stop."

Note: try to make sure that the one-, two-, or three-step instructions are the last thing the children hear before the music starts.

Researcher directions:
3. Make sure both cameras are still recording.

Appendix F:
Group-Based Sustained Attention Administration Protocol

<u>Supplies</u>
- Speakers
- Music CD/MP3 Player
- Protocol sheet
- Labels for children
- Class list for matching children and numbers

Marching Task:
During the Marching game, children stand in a circle facing the middle. They are instructed to march on one direction if one song begins playing (left for Happy), and the opposite direction if another song begins to play (right for Lego). After a few seconds, the song will stop and children will be directed to turn back and face the center before starting the next trial.
Researcher directions:
10. Get the children in a circle again and have researcher stand in center of circle.
11. Say, "Face toward me" and wait for students to face researcher.
12. Say, "We are going to play another game. You are going to hear one of two songs. When you hear this song (**Happy**), you will turn **left**, or this way (model by turning in circle), and start marching. When the music stops, stop marching and face me."
13. Say, "If you hear this song (**Lego**), you will turn **right**, or this way (model by turning in circle), and start marching. Remember, when the music stops, you stop marching and face me."
14. Say, "Okay, let's try!" (Practice)
15. Play **Lego**, and correct as needed. "Remember, when the music stops, stop marching and face me."
16. Play **Happy**, and correct as needed. Remember, when the music stops, stop marching and face me."

Researcher directions:
4. Make sure both cameras are still recording.
5. Control the music, letting it play for 7-15 seconds at a time.

Appendix G:
Group-Based Executive Function Coding Template

Response Inhibition						
Trial	Time	Stop Time (ms)	Distractibility	Freeze Time	Trial Time	
1						
2						
3						
4						
5						
Total:		0	0	0	0	

Working Memory Game						
Trial	Time	Stop Time (ms)	Distractibility	Action Recall	Action Performance	
1						
2						
3						
Total:		0	0	0	0	

Attention Control Game						
Trial	Time	March Direction Time (ms)	Stop Time (ms)	Distractibility	Performance (March Direction)	Recall
1						
2						
3						
4						
5						
6						
7						
Total:		0	0	0	0	0

Appendix H:
Counting Span Task Administration Protocol (Wechsler, 1991)

<u>Instructions Used with Numerical Memory Task (Forward Series)</u>

"Now let's see how well you can say numbers. Listen. Say 2- 6." (This practice is used as an introduction and will not be scored.) After the child responds, the experimenter will continue with the forward series, which contains six sets (Trial 1) by saying, "**Now say these numbers**." The experimenter will present the digits in the series one at a time, at the rate of one digit per second.

If the child successfully completes an item in Trial 1, the experimenter proceeds to the next item. If the child fails an item in Trial 1, the experimenter uses the item from Trial 2 which has the same item number as Trial 1. This allows the child another opportunity to successfully complete an item. If the child completes the item from Trial 2, the experimenter proceeds with the next item number in Trial 1. If the child fails an item from Trial 2 of Assessment 1, the experimenter goes on to the first item of Assessment 2.

After completing the two-forward series, proceed to the two-backward series with the following directions.

<u>Instructions Used with Numerical Memory Task (Backward Series)</u>

"Now I want you to say some more numbers. This time I want you to say them backwards. For example, if I say 3-5, you would say 5-3. Do you understand? What do you say when I say 7-2?"

If the child responds correctly, the experimenter continues with the backward series.

If the child does not respond correctly, the experimenter will say, **"No, you would say 2-7. I said 7-2. To say it backwards, you would say 2-7."** Then, experimenter will give one more backwards example, **"Now try this—what would you say if I said 4-1?"** If child gets it correct, go on… if not, explain the correct answer again, and then go on anyway…"**Now let's try some more."**

The experimenter will present the backward series in the same manner as the forward series, giving a second trial only if the first trial is failed.

Appendix I:
Counting Span Task Scoring Procedure (Wechsler, 1991)

Score: Correct/Incorrect

Assessment 1 (Forward Series)
Practice: "Say 2 - 6"

Length	Trial 1	Trial 2	Trial 3	Trial 4
2	5-8	4-9	6-1	2-5
3	6-9-2	5-8-3	2-7-4	4-3-9
4	3-8-1-4	6-1-8-5	4-3-9-6	1-7-6-8
5	4-1-6-9-2	9-4-1-8-3	5-2-1-7-4	8-5-9-3-2
6	5-2-9-6-1-4	8-5-2-9-4-6	9-3-1-7-5-4	4-7-6-2-1-3
7	8-6-3-5-2-9-1	5-3-8-2-1-9-6	1-2-4-8-3-5-9	7-9-4-3-2-5-8
8	3-6-1-8-4-2-7-5	7-9-1-3-5-4-6-2	4-8-3-7-1-2-6-5	3-5-7-6-9-4-8-1
9	1-4-7-5-8-3-2-9-6	9-1-6-8-2-7-4-5-3	8-4-2-7-9-1-5-6-3	6-8-1-7-4-5-3-2-9
Total				

Assessment 2 (Backward Series)
Practice: "If I say 3-5, you would say 5-3. What do you say when I say 7-2?"

Length	Trial 1	Trial 2	Trial 3	Trial 4
2	9-6	4-1	6-2	1-5
3	1-8-3	2-5-8	3-7-4	7-3-2
4	5-2-4-9	6-1-8-3	9-1-6-4	8-5-2-6
5	1-6-3-8-5	6-9-5-2-8	7-3-5-8-2	4-7-1-9-5
6	4-9-6-2-1-5	3-8-1-6-2-9	5-9-4-6-8-3	1-8-2-6-3-9
Total				

Appendix J:
Woodcock Johnson III Tests of Cognitive Ability
(Woodcock, McGrew, & Mather, 2001; 2007)

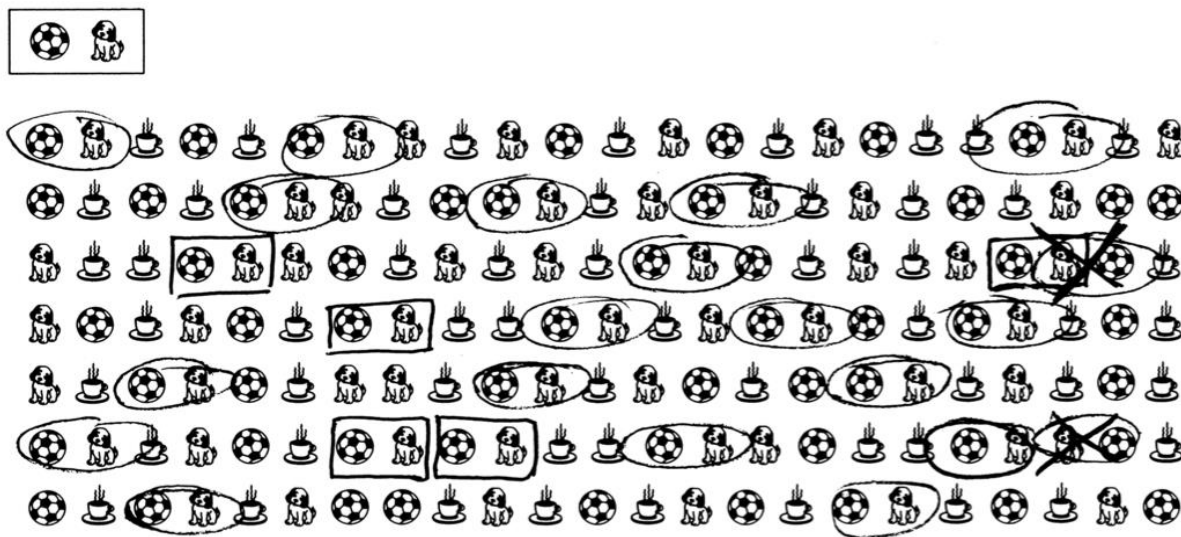| Test	Area/Narrow Ability	Cognitive Process(es)	Related Educational Interventions
Test 20: Pair Cancellation	Processing Speed *Reading decoding*	Controlled, focal attention; vigilance	Speed drills; repetition

FIGURE 10.1 This sample from the *Pair Cancellation test* (Woodcock-Johnson III Tests of Cognitive Abilities; Woodcock, McGrew, and Mather, 2001c) shows how scanning cancellation tests with horizontally aligned stimuli can elicit subtle unilateral inattention—usually on the left. These top seven (of 21) lines contain four of the eight left-sided omissions (enclosed in rectangles), one of the three right-sided omissions, and two right-sided errors (X'd) made by the 55-year-old dermatologist who had sustained a blow to the left side of his head in a skiing accident (see p. 87). (© Riverside Press. Reprinted with permission) From Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). Neuropsychological Assessment, Fifth Edition. New York, NY: Oxford University Press.

Woodcock Johnson III Tests of Achievement
(Woodcock, McGrew, & Mather, 2001; 2007)

Test	Area/Narrow Ability	Cognitive Process(es)	Related Educational Interventions
Test 1: Letter-Word Identification	Reading *Reading decoding*	Feature detection and analysis (for letters) and recognition of visual word forms and/ or phonological access to pronunciations associated with visual word forms (i.e., words may or may not be familiar)	Explicit, systematic; synthetic phonics instruction; word-recognition strategies (word walls, flow lists, word banks, flash cards); repeated readings, teaching high-frequency words, spelling-based decoding strategies; Fernald method
Test 9: Passage Comprehension	Reading *Reading comprehension Cloze ability*	Construction of propositional representations; integration of syntactic and semantic properties of printed words and sentences into a representation of the whole passage; inferential bridging	Vocabulary enrichment; activating prior knowledge; use of graphic organizers; self-monitoring strategies; memory and imagery strategies; *Multipass*
Test 10: Applied Problems	Mathematics *Quantitative reasoning Math achievement Math knowledge*	Construction of mental mathematics models via language comprehension, application of math knowledge, calculation skills, and/or quantitative reasoning; formation of insight	Use of pictures and diagrams; direct instruction; use of data-tables; strategy instruction

Appendix L:
WJ-III Description and Administration Protocol

A. General Description

The Woodcock-Johnson is a standardized achievement measure that contains subscales that measure a variety of content areas. For these assessment points, we will use six of the subscales: letter-word identification, reading fluency, passage comprehension, calculation, math fluency, and applied problems.

Reference:
Woodcock, R. W., McGrew, K., & Mather, N. (2001). The Woodcock-Johnson Tests of Achievement: Third edition. Itasca, IL: Riverside.

B. General Task Preparation

This task will be **videotaped**.

C. General Task Procedure

Two of the tests, Letter-Word Identification and Calculation, will be scored by the experimenter on the Test Record form. For each item, mark 1 for correct and 0 for incorrect. The remaining two tests, Reading and Math Fluency, will be scored by the child's responses in the worksheets at a later time. Be sure to have the Test Record forms, the child worksheets, and a timer ready to go.

Introduce the tests by saying:
        "**I am going to ask you to answer some questions and to solve some problems. These questions and problems can be used with children of all ages and even adults, so some of them will be easy, and others will seem hard. Just try to do your best.**"

I. Letter-Word Identification- Preparation and Procedure

        The recommended starting point for a Kindergarten student is item 1. However, remember that before moving forward, you must first establish a basal (6 lowest items in a row correct). If a child misses one of the first 6 items administered, complete the page and then move backward one page of words at a time until a basal has been established. The test

can be stopped when a child reaches a ceiling (6 items in a row incorrect). Once a child misses 6 items in a row, complete the page and then move on to the next test. Even if the child accurately reads a word above the ceiling while completing the page, the test can still be discontinued.

To begin the test:

Place the test easel in front of the child, point to the first word (**#1**, p.), and say **"What is this word?"** After the child responds, say **"Go ahead with the others. Don't go too fast."**

You can point to each word if it is necessary to get a response. If the child does not respond within 5 seconds, point to the next word and continue. Score all administered items on the Test Record form with either 0 (incorrect) or 1 (correct). Be sure that you know the correct pronunciation for each word, and do not help the child read any words. If you can not clearly hear a response, let the child finish the page and then have him or her repeat the word in question. Rescore that item appropriately. If the child pronounces words letter by letter or syllable by syllable, say **"First read the word silently and then say the word smoothly."** Give this instruction only once, and words that are not read smoothly are scored as incorrect.

II. Reading Fluency- Preparation and Procedure

This test has a 3-minute limit. Use the timer to be sure that the child has exactly 3 minutes. If he or she stops early, record the *exact* amount of time used on the Test Record form so that scoring can be adjusted.

Practice items:

Begin with the Reading Fluency sample items on the child worksheet. Place the worksheet in front of the child and say, **"I want you to read some sentences and decide if the answer is yes or no."**

Point to Sample Item A and say, **"Look at this sentence. It says, "A cow is an animal."** (pause) **"Is that true?"** (Pause for response.) **"Because the answer is yes, you would circle the letter Y"** (Point to circled "Y") **"Now look at the second sentence. It says, "A fish lives on land."** (pause) **"Is that true?"** (Pause for response.) **"Because the answer is no, you would circle the letter N"** (Point to circled "N".)

Now give the child the pencil and say **"Now look at the next four sentences. Draw a circle around the correct answer for each sentence. Work as fast as you can without making mistakes. Go ahead."**

IF THE CHILD MAKES AN ERROR on the practice items, score the practice item with a 0. Then say, **"Read the sentence aloud and tell me if the answer is yes or no."** If the child is still incorrect, explain the sentence and the correct answer.

IF THE CHILD DOES NOT RESPOND to an item, score that item with a 0. Then say, **"Read the sentence aloud and tell me if the answer is yes or no."** If the child cannot read the sentence, point to the next one and say, **"Try the next one."**

IF THE CHILD GETS FEWER THAN 3 CORRECT on the practice items, discontinue testing and record a score of 0 on the Test Record form.

To begin the test:

Open the packet to the beginning of the Reading Fluency items and hold it up away from the child's view. Say, **"Start here** (point to first sentence) **and read as many sentences as you can. Decide if the answer is yes or no. After you get to the bottom** (point to the bottom of the first column)**, go to the top** (point to the top of the second column). **There are three pages. Keep working until I tell you to stop. Work as fast as you can without making mistakes. If you do make a mistake, cross out the one you do not want. If you have trouble reading a word or cannot think of the answer, skip that one and go on to the next one. You will have three minutes. Tell me if you finish before I say, 'Stop'."**

Place the packet in front of the child and say, **"Go ahead."** Begin timing 3 minutes. Make sure the child continues to the top of the next page after finishing each page. After 3 minutes say, **"Stop. Put your pencil down."**

If the child finishes in less than 3 minutes or refuses to continue, record the exact amount of time used on the Test Record form.

III. Passage Comprehension - Preparation and Procedure

All children will begin this test on Sample Item B. After administering the practice item, it is recommended that a fifth grade student start with item 20. However, remember that before moving forward from the starting point, you must first establish a basal (6 lowest items in a row correct). If a child misses one of the first 6 items administered, complete the page and then move backward one page at a time until a basal has been established. The test can be stopped when a child reaches a ceiling (6 items in a row incorrect). Once a child misses 6 items in a row, complete the page and then tell the child that he or she is finished. Even if the child accurately answers an item above the ceiling while completing the page, the test can still be discontinued.

Practice item:

Place the test easel in front of the child, opened to Sample Item B (p. 155). Point to the picture on the child's side and say, "**Look at this picture. Listen. This says, 'The house is bigger than the…**' (Pause) **What word belongs in the blank space**?"

If the child responds with one of the words listed as correct, continue on to the item listed as a starting point. Otherwise, if the response is incorrect or the child does not respond at all, give him or her the correct answer. Then repeat the practice item and allow the child to respond before continuing.

Administering test items:

Point to **item 20** (p.163) on the page and say, **"Read this to yourself and tell me one word that goes in the blank space."** (Point to the blank.) The passages should be read silently. If the child is reading aloud, ask him or her to read silently. However, if he or she continues to read aloud, do not insist on silent reading. Score all administered items on the Test Record form with either 0 (incorrect) or 1 (correct). If a response is incorrect, write the child's response next to that item. Remember that a basal of 6 lowest items correct must be established. The directions do not need to be given for more than the first item as long as the child understands the task and continues.

If the child does not respond within 30 seconds after reading the whole passage, encourage him or her to respond. Then move to the next question, point, and say, **"Try this one."**

Only items listed on the "Correct" line can be given credit. Acceptable variations include a different verb tense or number (singular/plural), unless the scoring guide indicates differently. If the child uses a different part of speech, such as a noun for a verb, score the item as *incorrect*. Occasionally a question will have an answer that requires a follow-up question, listed as "Query". If the child responds with a word listed next to "Query", say **"Tell me another word."** Unless noted, only one-word answers are acceptable. If the child responds with two or more words, ask him or her for a one-word answer.

Do not, for any reason, help the child to read any words on the test.

IV. Calculation- Preparation and Procedure

Have every child start on number 9. Remember that a basal (6 lowest items in a row correct) must still be established. If a child misses one of the first 6 items administered, have him/her go back and complete the previous row in order to establish a basal. The experimenter must score each item in the Test Record form as the child works in order to know whether the basal has been established and when a ceiling has been reached. The test is discontinued when the child incorrectly answers or skips 6 items in a row. Skipped items are scored as incorrect. Score poorly formed or reversed numbers as correct.

To begin the test:

Get out the packet with the Calculation test items and make sure the child has a sharpened pencil. Say, **"I want you to do some math."** Point to **Item 9** and say, **"Begin with number nine and answer as many problems as you can. If you come to one that you do not know how to do, just skip it and try the next one."** Do not read items or offer assistance in any other way.

If the child is having difficulty with a problem, encourage an attempt but do not allow him or her to spend unnecessary time. Fraction answers should be simplified. If the child writes 2/4, say **"Simplify your**

answer." Once 6 problems in a row have been skipped or missed, proceed to the next test.

V. Math Fluency- Preparation and Procedure

This test has a 3-minute limit. Use the timer to be sure that the child has exactly three minutes. If he or she stops early, record the *exact* amount of time used on the Test Record form so that scoring can be adjusted. All children will begin with the first problem.

To begin the test:

Get out the packet with the Math Fluency test items and hold the problems away from the child's view. Say, **"I want you to work some simple arithmetic problems. Start here.** (Run your finger across the first row.) **When you finish a row, go to the next one and work each problem until you finish the page. Then go to the top of the next page. If you cannot think of an answer, skip that item and move to the next one. Work as fast as you can without making mistakes. Be sure to watch the signs. If you do make a mistake, just cross out the answer you do not want. You will have 3 minutes. Tell me if you finish before I say, 'Stop'."**

Place the test items in front of the child and say, **"Go ahead."** Begin timing 3 minutes. Do not remind him or her about signs during the test. Make sure he or she continues to the next page after finishing the first. After 3 minutes say, **"Stop. Put your pencil down."**

If the child finishes in less than 3 minutes or refuses to continue, record the exact amount of time used on the Test Record form.

VI. Applied Problems - Preparation and Procedure

Remember that before moving forward from the starting point, you must first establish a basal (6 lowest items in a row correct). If a child misses one of the first 6 items administered, complete the page and then move backward one page at a time until a basal has been established. The test can be stopped when a child reaches a ceiling (6 items in a row incorrect). Once a child misses 6 items in a row, complete the page and

then tell the child that he or she is finished. Even if the child accurately answers an item above the ceiling while completing the page, the test can still be discontinued.

To begin the test:

Begin with **item 25** (p.195). Place the test easel in front of the child with the numbered questions facing you. All instructions and correct answers will be facing you. Read everything in blue for the question, and point as directed. For some questions, there will be child answers that require you to further prompt the child. These answers and the required prompt will be listed beside the word "Query" under the question. Score all correct responses (including correct responses following a query) with a 1 on the Test Record form. Score incorrect or lack of responses with a 0 on the Test Record form.

Beginning with item 30, provide the child with paper and a pencil. Say, **"You can use this paper and pencil if you need them."** Paper and pencil can be provided earlier if requested by the child. You may repeat questions whenever the child asks.

If the child is taking an unnecessarily long time to answer a question, encourage a response. If he or she does not respond within a reasonable amount of time, point to the next question and say, **"Let's try this one."**

Appendix M:
Teacher Reported Child Executive Function Scale
Children's Behavior Questionnaire (Putnam, & Rothbart, 2006)

Question	Label	Factor
When practicing an activity, has a hard time keeping her/his mind on it	CBQ 1	Attention control
Will move from one task to another without completing any of them	CBQ 2	Attention control
When drawing or coloring in a book, shows strong concentration	CBQ 3	Attention control
When building or putting something together, becomes very involved in what s/he is doing, and works for long periods	CBQ 4	Attention control
Is easily distracted when listening to a story	CBQ 5	Attention control
Can wait before entering into new activities if s/he is asked to	CBQ 7	Response Inhibition
Has trouble sitting still when s/he is told to	CBQ 9	Response Inhibition
Can easily stop an activity when s/he is told "no"	CBQ 12	Response Inhibition

Appendix N:
Teacher Reported Working Memory Scale

Question	Label	Factor
Loses track during complicated tasks and may eventually abandon these tasks	Original Item	Working Memory
Makes place-keeping errors (e.g., skipping or repeating steps)	Original Item	Working Memory
Shows incomplete recall of information	Original Item	Working Memory