THE RELATIONSHIP BETWEEN EXECUTIVE FUNCTIONING AND EXTERNALIZING BEHAVIOR PROBLEMS DURING ELEMENTARY SCHOOL

BY

ZACHARY T. LONGO

A DISSERTATION SUBMITTED TO THE FACULTY OF

ALFRED UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PSYCHOLOGY

IN

SCHOOL PSYCHOLOGY

ALFRED, NEW YORK

2019

THE RELATIONSHIP BETWEEN EXECUTIVE FUNCTIONING AND EXTERNALIZING BEHAVIOR PROBLEMS DURING ELEMENTARY SCHOOL

BY

ZACHARY T. LONGO

SUNY CORTLAND, B.A. (2015)

ALFRED UNIVERSITY, M.A. (2017)

AUTHOR Zachary T. Longo

APPROVED BY Lynn O'Connell, Psy.D.
Committee Chairperson

Amy Button, Psy.D.
Committee Member

Steve Byrne, Psy.D.
Committee Member

ACCEPTED BY Andrea Burch, Psy.D.
Program Director, School Psychology Doctoral Program

ACCEPTED BY Kevin Curtin, Ph.D.
Chair, Division of Counseling, School Psychology, & College Student Development

ACCEPTED BY John D. Cerio, Ph.D.
Dean, School of Graduate and Continuing Studies

ACCEPTED BY W. Richard Stephens, Ph.D.
Provost and Vice President for Academic Affairs

Acknowledgements

The dissertation. Two words that strike fear in any doctoral student wishing to one day know what it feels like to be done with school. To say this process didn’t leave me feeling like Sisyphus at times would be a lie. But here I stand – atop the mountain – and I can assure you that the feeling of freedom and accomplishment was worth every hour. As with many great accomplishments though, this reward is not just for the person holding the trophy at the end. I owe it to many individuals who graciously gave me a portion of their lives to help me succeed. From long discussions aimed at improving my skills as a researcher, to simply listening about the intricacies of factor analysis, many individuals embodied benevolence and understanding in order to help me achieve this lofty goal.

To my chair, Dr. O’Connell, your words of encouragement, your expertise, and your authenticity served as a power surge through the many nights I spent puzzled and tired. To my committee, Dr. Byrne and Dr. Button, your unique perspectives constantly challenged me, forcing me to grow intellectually and enhancing this project immensely. To Dr. Greil, one of the only people who enjoys talking statistics, a title that comes far short of the incredible knowledge base you have secured through a lifetime as a successful researcher and professor – I thank you for the hours you spent working through the intricacies of complex survey design and structural equation modeling with me, despite being in retirement. To Dr. J. Atlas, your meticulous feedback purged the early forms of this document from many conceptual and grammatical flaws.

To my cohort, and the rest of the Alfred University School Psychology program as well as those who have retired, one would have a difficult time finding individuals who are more devoted to producing high-quality work and helping others do so as well. To my family and friends, I thank you for never punishing me for the time I devoted to this project. Your

willingness to listen, offer praise, and a place of respite has not gone unnoticed. This project is a manifestation of the incredible support network you provided and I only hope to reciprocate this powerful gift in the years to come.

I'll close with some additional words for my parents, Thomas and Lynne Longo. I bet I gave you more than a few scares throughout my childhood/adolescence. From teacher complaints about my behavior in elementary school to traveling the country for baseball, oh what a long strange trip it's been. The completion of this project is a testament to the environment you provided me, the incredible sacrifices you made, and the love I received and continue to receive to this day. I dedicate this project to you all and hope that I can contribute to your own goals in the way you helped me achieve mine.

TABLE OF CONTENTS

I. INTRODUCTION 1

a. Components of Executive Function..2

b. Development of Executive Functioning 2

c. Correlates of Executive Functioning 3

d. Current Study..7

e. Hypotheses..8

II. LITERATURE REVIEW 10

a. Executive Functioning in Context 10

b. Current Models of Executive Function..11

c. An Emerging Conceptualization of Executive Functioning 14

i. The role of Inhibition...15

d. The Relationship between Executive Function
and Other Domains of Functioning..16

i. Relationship between executive function and academic achievement 16

ii. Relationship between executive function and social-emotional
functioning 17

1. Executive function and internalizing disorders..............................18

2. Executive function and externalizing behavior problems.............19

e. The Relationship between Executive Function and Demographic Variables........21

i. Race and executive functioning 22

ii. Sex and executive functioning..23

iii. Socioeconomic status and executive functioning 26

f. Developmental Trajectory of Executive Functioning's Components....................27

i. Working Memory..28

ii. Cognitive Flexibility..32

iii. Inhibition..35

g. Current Study...38

III. METHOD...41

a. Database Overview...41

i. Participants..41

b. Variables...42

i. Control variables..43

1. Sex...43

2. Race...44

3. Socioeconomic status..45

ii. Independent variables...45

1. Numbers Reversed...46

a. Reliability..47

b. Validity...48

2. Dimensional Change Card Sort...51

a. Reliability..55

b. Validity...56

iii. Dependent variable...57

1. Reliability..58

2. Validity...59

c. Data Preparation......60
i. Weighting and jackknife variance estimation......60
ii. Missing data......61
d. Analyses......62
i. Model......64
IV. RESULTS......67
a. Analysis Summary......68
b. Findings Related to Control Variables......69
i. Sex......70
ii. Race......70
iii. Socioeconomic status......72
c. Main Findings Related to Hypotheses......73
i. Hypothesis 1......73
ii. Hypothesis 2......74
iii. Hypothesis 3......74
iv. Hypothesis 4......75
v. Hypothesis 5......76
d. Findings Related to Standardized Covariances......77
e. Working Memory's Contribution to Cognitive Flexibility......78
V. DISCUSSION......80
a. Implications of Findings......82
i. Development of executive functioning......82
ii. Executive functioning's relationship to EBP......83

b. School-Based Implications 87

c. Limitations 89

d. Future Research 92

e. Final Remarks 95

VI. REFERENCES 97

VII. Figure 1 Theoretical Model 114

VIII. Figure 2 Hypothesized Model..115

IX. Figure 3 Model Depiction of Main Hypotheses 116

X. Figure 4 Model Depiction of Standardized Covariances for Control Variables........117

XI. Figure 5 Standardized Covariances among Independent and Dependent Variables .118

XII. Results of Main Hypotheses 119

XIII. Table 1 Sample Characteristics...120

XIV. Table 2 Descriptive Statistics of Variables 121

XV. Table 3 Summary of Direct, Indirect, and Total Effects for Externalizing Behavior Problems 122

XVI. Table 4 Summary of Direct, Indirect, and Total Effects for Working Memory........123

XVII. Table 5 Summary of Direct, Indirect, and Total Effects for Cognitive Flexibility ...124

Abstract

The present study investigated the relationship between two major components of executive functioning (EF) and externalizing behavior problems (EBP) during the early elementary years. More specifically, Working Memory (WM) and Cognitive Flexibility (CF) measured in kindergarten through second grade were used to predict teacher ratings of EBP in the spring of second grade utilizing the Early Childhood Longitudinal Study: Kindergarten class of 2010-2011 (ECLS-K) dataset. Path analysis was implemented in order to understand the predictive relationship of WM and CF on EBP, as well as to uncover the developmental influence of WM on CF. After controlling for race, sex, and socioeconomic status (SES), results indicated that lower levels of EF predicted a higher magnitude of EBP, as measured by teacher rating. Specifically, of all the time points for WM and CF, only fall of kindergarten CF and fall of second grade WM significantly predicted EBP, but only to a small degree. Developmentally, WM was also found to predict later levels of CF at numerous timepoints, providing support to the notion that successful CF is built upon development of WM as one cannot flexibly shift between perspectives until one can hold a perspective in present awareness first.

Chapter I: Introduction

Executive functions represent an array of cognitive processes that dictate the degree to which individuals can readily adapt to their environment, allowing for the selection of behaviors that achieve specific goals ranging from learning in the classroom to successfully interacting with peers. To investigate executive functioning (EF), researchers (e.g., Best, Miller, & Naglieri, 2011; Schoemaker, Mulder, Deković, & Matthys, 2013) have frequently used a model that was initially proposed by Miyake and colleagues (Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000). In this seminal article concerning the organization of EF, Miyake and colleagues demonstrated via confirmatory factor analysis (CFA), that EF is best organized in a *Unity/Diversity Framework*. The *Unity/Diversity Framework* postulates that EF is represented by three factors that uniquely contribute to an overarching, shared, higher-order construct identified as Common EF. The factors of Common EF include Inhibition, Working Memory (WM), and Cognitive Flexibility (CF).

Similarities between the *Unity/Diversity Framework* of EF and one of the dominant theories of intelligence, the Cattell-Horn-Carroll (CHC) model of intelligence (McGrew, 2009), are easily identified. In CHC theory, intelligence is represented by a three-stratum hierarchy with overall intelligence represented as *g*. Underneath *g*, various broad abilities (e.g., Fluid Reasoning, Comprehension-Knowledge, Short-Term Memory, Long-Term Storage Retrieval, Visual Processing, & Auditory Processing, etc.) exist which are then further broken down into another stratum of various narrow abilities (McGrew, 2009). As with CHC theory, the *Unity/Diversity Framework* represents Overall EF as the highest-order factor, with narrower components (i.e., Inhibition, WM, and CF) represented at a lower stratum. The basic essence of these models is that the constructs they purport to describe are multi-faceted in that many unique components come together to represent a shared phenomenon (e.g., intelligence or EF).

Components of Executive Functioning

The unique components of the *Unity/Diversity framework* were selected by Miyake and colleagues (2000) based on what they found in the literature to be the three most frequently discussed components of EF: Inhibition, WM, and CF. Inhibition is defined as the suppression of a dominant response. In other words, a child would be exhibiting Inhibition if he or she could prevent themselves from producing a behavior in a situation that had a powerful pull towards that behavior (e.g., electing not to eat a treat that is sitting in front of him/her). Working Memory, otherwise known as *Updating*, is generally defined as one's ability to process incoming sensory information and manipulate it in present awareness. For example, a child would be exhibiting WM when listening to teacher directions and rehearsing each step of the directions in his/her head while carrying out the steps. Finally, CF, otherwise known as *Shifting*, is defined as one's ability to inhibit a dominant response to a particular stimulus in favor of an alternate response that produces a desirable outcome. For example, a child would be exhibiting CF if they were able to prevent themselves from emotional dysregulation in response to something aversive, while simultaneously choosing an alternate more adaptive behavior.

Development of Executive Functioning

Developmental research concerning how each of these components matures over time has provided some insight into each component's developmental timeline although much remains to be answered. In their review of the literature, Best and Miller (2010) utilized the *Unity/Diversity Framework* of EF to understand how Inhibition, WM, and CF mature independently over time. In summary, the researchers found that all three components of EF exhibit a prolonged development, starting as early as infancy and lasting all the way through adulthood. However, some variation in the developmental timeline among the EF components does exist. For example, Inhibition and Working Memory are thought to begin developing in infancy whereas CF is thought to develop later, beginning around three to four years old; (Best & Miller, 2010; Best, Miller, & Jones, 2009) as CF is seemingly built upon the development of WM and Inhibition. A review of the definitions of these constructs provides explanation as to why this might be the case. In particular, CF necessitates that the individual not only processes incoming information via WM, but then must inhibit a dominant response via Inhibition, and then produce an alternate response that produces a favorable outcome via CF. Consequently, CF has been theorized to develop following the onset of Inhibition and WM (Best et al., 2009; Blakey, Visser, & Carroll, 2016). While there is a developing theoretical understanding of the relationship between Inhibition, WM, and CF, more research is needed to quantify the contribution that either Inhibition or WM make to the development of CF. But what makes studying the development of EF important in the first place?

Correlates of Executive Functioning

Historically, researchers and practitioners, alike, have been interested in the relationship between EF and academic achievement. In like fashion, research has found that intelligence, or cognitive abilities, are somewhat related to EF and we know that cognitive abilities are predictive of learning behaviors (e.g., Blair, 2006; Friedman, Miyake, Corley, Young, DeFries, & Hewitt, 2006). Thus, to what degree does EF influence one's ability to learn? Deficits in EF during childhood manifest in an increased probability for academic issues ranging from difficulties in reading and math skill development to difficulties with social-emotional functioning in the academic setting (e.g., Bock, Gallaway, & Hund, 2015; Ciairano, Bonino, &

Miceli, 2006; Morgan, Hui, Farkas, Cook, Hung Pun, & Hillemeier, 2017). This latter point related to the relationship between EF and social-emotional development is only beginning to be studied.

In the academic environment, the child is tested with a variety of social experiences that can have significant implications for their success in school. For example, a child's skills in the play environment can impact the development of friendships with peers. Conceptually, the link between EF and successful navigation of social environments is clear; the ability to process incoming social information, regulate emotional arousal and inhibit maladaptive behaviors, and then select an appropriate behavior has logical connections to social success at all ages.

Although in its infancy, the current literature base surrounding the relationship between EF and social-emotional functioning suggests an important link. For example, a meta-analysis investigating the relationship between EF and antisocial behavior demonstrated that groups of participants with antisocial characteristics (e.g., criminality, externalizing behavior disorder, & antisocial personality disorder) performed significantly lower on measures of EF relative to healthy controls, finding a small to medium effect size ($d = 0.44$; Ogilvie, Stewart, Chan, & Shum, 2011). Similarly, in their review of the literature, Snyder, Miyake, and Hankin (2015) found that impairments in EF manifested in difficulties such as ruminating, worrying, as well as less frequent use of adaptive emotional regulation strategies, such as reappraisal. Further, in the classroom, Ciairano et al. (2006) demonstrated that children high in a particular aspect of EF, Cognitive Flexibility, were more likely to exhibit increased cooperative social behaviors, such as helping a classmate with a puzzle.

One aspect of social-emotional functioning, the presence of externalizing behavior problems (EBP), has clear connections to the regulatory function of EF, as the characteristics of children with EBP mirror an individual struggling with self-regulation. In particular, EBP are generally defined as behaviors such as aggression, impulsivity, defiance, disruptiveness, and hyperactivity (Hinshaw, 1992). Behavioral dysregulation occurs when an individual is unable to modulate his/her behavior to fit within the conventionally acceptable range in a given setting. For example, a student may be unable to sit still during quiet reading time because this is an undesirable task for them. Instead of inhibiting their negative emotions toward the task, the student may exhibit defiant or distracting behaviors despite environmental cues to stop (e.g., teacher telling the student to sit quietly).

The current literature provides evidence supporting this logical connection between EF and EBP. Schoemaker et al. (2013) conducted a meta-analysis utilizing the *Unity/Diversity Framework* to understand the unique relationship between the three constructs of EF (i.e., Inhibition, WM, and CF) and EBP in preschoolers. The inclusionary criteria for articles included: (1) a sample of children with EBP as defined by a clinical diagnosis of Attention-Deficit Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD)/Conduct Disorder (CD), or aggressive or hard to manage children, as measured by formal tests or semi-structured interviews; (2) an average age of children between 3.0 years and 6.0 years; and (3) measures of EF, including specific tests that tap into Inhibition, WM, and/or CF.

The results of this salient meta-analysis indicated a medium effect size between overall EF and EBP ($r = 0.22$), a medium effect size between Inhibition and EBP ($r = 0.24$), and small effect sizes between both WM and CF and EBP ($r = 0.17$ & $r = 0.13$, respectively). These findings led the researchers to conclude that EBP are related to impairments in EF even during early childhood as operationally defined as the preschool years (i.e., three years through six years). Further, their findings illustrate that, at least to some degree, unique components of EF such as WM and CF may differentially relate to the presence of EBP.

Another example of the link between EF and EBP comes from research conducted by Sulik, Blair, Mills-Koonce, Berry, and Greenberg (2015) which found that EF longitudinally mediated between parenting practices and EBP using a sample of 1,115 low socioeconomic status (SES), rural children ages 36 months to 90 months. Their results suggested that sensitive parenting practices, as determined through expert video coding of parent-child interactions, influenced fewer EBP primarily through the fostering of EF within the child. Specifically, no significant direct effects between parenting practice and EBP were observed whereas significant direct effects between EF and EBP were found. Thus, sensitive parenting practices alone did not influence the presence of EBP nearly as much as when parents taught children to use Executive Functioning.

In summary, prior investigations have demonstrated a relationship between EF and social-emotional functioning. Despite this developing body of research, many questions still remain unanswered. For example, much of the current research available utilizes general measures of EF rather than distinct factors of EF. In addition, many studies lack a consistent theoretical foundation. Thus, it is difficult to generalize across studies as the measures of EF are not the same nor are they purporting to measure the same factors of EF. Further, few studies have investigated how unique components of EF may differentially relate to social-emotional functioning, as well as how they relate to each other. For example, we currently do not fully understand the development of EF and its components, nor do we understand the link between the individual components of EF and EBP.

Current Study

As stated earlier, previous studies have demonstrated a link between EF and EBP (e.g., Schoemaker et al., 2013; Sulik et al., 2015); however, no study to date has utilized path analysis to simultaneously measure the development of unique components of EF and how they differentially relate to EBP. The current study aimed to address this gap in the literature. The goal was to examine the longitudinal relationship between two major components of EF (i.e., WM and CF) and the presence of EBP. Please see Figure 1 and Figure 2 for a graphic depiction. Unfortunately, Inhibition was not analyzed as the current study used an extant dataset in which no measure of Inhibition was administered. Despite this weakness, the current study hoped to provide valuable information concerning the other two factors of EF and their role in EBP, while simultaneously advancing the developmental literature on EF as well. Regarding the latter, this study also examined the manner in which WM plays on the development of CF; an area that needs further exploration in order to better understand how each component of EF unfolds over the lifetime. Using the Early Childhood Longitudinal Study: Kindergarten class of 2010-2011 (ECLS-K) dataset (Tourangeau et al., 2017), the current study yielded highly generalizable findings, as this large sample of children was nationally representative of kindergarteners in the U.S. population concerning numerous key background variables (e.g., SES, race).

Notably, the longitudinal design allowed the variables to be temporally ordered, allowing for stronger assumptions of causality, as later measures of a variable cannot influence earlier measures of another variable. These findings have direct implications for not only theory concerning EF, but also clinical practice. Regarding theory, the current study uncovers how CF develops in relation to WM, allowing future researchers to better understand whether a developmental hierarchy is present. Finally, clinical practice is hopefully improved by providing evidence as to the robustness of the relationship between specific components of EF (i.e., WM and CF) and EBP. This furthers assessment practices by providing preliminary evidence of the predictive validity of specific components of EF. Further, clinicians may be able to use these measures to identify students in need of interventions targeted at improving the child's EF.

Hypotheses

Based on prior empirical investigations from the EF literature base (e.g., Schoemaker et al., 2013; Sulik et al., 2015) as well as the logical relationship between EF and EBP, the main hypotheses of the present study are:

(1) WM in the fall of kindergarten will have a significant negative relationship with EBP in the spring of second grade;

(2) CF in the fall of kindergarten will have a significant negative relationship with EBP in the spring of second grade;

(3) CF in the spring of kindergarten will mediate, to a significant degree, the relationship between WM in the fall of kindergarten and EBP in the spring of second grade;

Please refer to Figure 3 for a visual depiction of the main hypotheses in path format. In addition to the three main hypotheses presented above, two exploratory hypotheses will also be investigated:

(4) WM in the fall of kindergarten will significantly predict CF in the spring of kindergarten, as the development of WM is a hypothesized prerequisite for the development of CF;

(5) The relationship between each component of EF (i.e., WM and CF) and EBP will grow stronger as the temporal distance between the measures of EF and later EBP grow closer. In other words, WM and CF in the fall of second grade will have a stronger relationship

with EBP, as measure in the spring of second grade, than WM and CF in the fall of kindergarten.

Chapter II: Literature Review

In general, executive functions are higher-order cognitive processes that allow individuals to regulate their behavior in everyday life. In their historical account of executive functioning (EF), Goldstein, Naglieri, Princiotta, and Otero (2014) found over 30 different definitions of EF that have been used in the literature from as early as 1966, with all definitions converging upon goal-directed behavior in some form. Beginning as early as infancy, humans rely on EF to successfully navigate their complex milieu, choosing to enact and inhibit certain behaviors in favor of adaptive outcomes (e.g., make a friend, stop crying, plan for an event, etc.). Failure to exhibit EF, particularly during childhood, can lead to negative outcomes that can unfold into significant issues later on, such as an inability to select appropriate behaviors for different settings or the development of mental illness.

Executive Functioning in Context

Individuals low in EF struggle to regulate emotions, follow multi-step directions, maintain attention on cognitively demanding tasks, and ignore distractions. These behavioral deficiencies leave the individual ill-equipped to pursue goals in occupational or school settings. Further, research has now associated deficits in EF to numerous developmental disorders such as ADHD, Autism Spectrum Disorder, Tourette's Syndrome, Obsessive-Compulsive Disorder, and Schizophrenia (Barkley, 2014; Johnson, 2012). In conjunction with established associations with various mental illnesses, EF deficits have also been associated with the presence of antisocial behavior as well as reading and mathematics difficulties (Best et al., 2011; Morgan et al., 2017; Ogilvie et al., 2011). Taken together, it is clear that individuals low in EF struggle in numerous ways.

In contrast to individuals low in EF, individuals high in EF are better able to plan, maintain attention levels, remember and apply information, and flexibly shift between different perspectives in order to solve a problem (Morgan et al., 2017). Research is also now suggesting that EF is predictive of social competence as early as the preschool years (Bock et al., 2015; Ciairano et al., 2006). These socially adaptive behaviors are crucial to success at all ages in virtually any setting. Finally, researchers are beginning to suggest that individuals with strong EF may be better able to cope with genetic and environmental risks to mental illness (Johnson, 2012). For those susceptible to the development of mental illness, the fostering of strong EF may be the key to avoiding the unfolding of mental illness later in life. For example, high EF may help the individual approach situations more flexibly than one with low EF who may be more apt to adopt a rigid thinking style that leads to maladaptive behavior.

Current Models of Executive Functioning

The literature on EF consistently identifies EF as a complex construct that involves an array of intricate neurological processes that underlie human behavior in important ways. A seminal study conducted by Miyake et al. (2000) suggested a conceptual framework that has informed numerous investigations of EF in both adults and children. Using a sample of 137 undergraduates from the University of Colorado at Boulder, their team conducted a CFA in order to resolve the longstanding controversy between two widely known frameworks of EF: the unitary vs. diversity theories of EF.

Those in favor of a unitary model of EF argued that EF is best represented as a single mechanism that reveals itself through different functions such as attention, emotional regulation, flexibility, inhibition, and WM. Support for this viewpoint relied primarily on frontal lobe dysfunction studies that revealed some common behavioral trends (e.g., impulsivity and disinhibition) among patients with damage specific to the frontal lobe. In contrast, those in favor of the diversity of EF argued for a fractionated account of EF, meaning that EF was thought to be composed of numerous separable, distinct components that each uniquely contributes to an overarching EF. This viewpoint was based upon numerous individual-difference studies that yielded small correlations ($r = .4 >$) between measures of EF, which has been a consistent finding within the EF literature (Friedman & Miyake, 2017; Miyake et al., 2000). For example, Lehto (1996) found no significant correlations between three widely-known measures of EF (i.e., The Wisconsin Card Sorting Test, the Tower of Hanoi, and the Goal Search Task) using a population of 15 to 16-year-old students.

To explore the *Unity/Diversity* theoretical framework of EF, Miyake et al. (2000) first identified three commonly postulated factors of EF and they are as follows: *Inhibition*, *Updating* (Working Memory [WM]), and *Shifting* (Cognitive Flexibility [CF]). Inhibition was defined as "one's ability to deliberately inhibit dominant, automatic, or prepotent responses when necessary (Miyake et al., 2000, p. 57)." Updating, or WM, was defined as the ability to process, hold, and manipulate incoming information by filtering out irrelevant information. Finally, Shifting was defined as the ability to switch between different mental rules or sets for solving problems. In order to reduce measurement error, three widely-used measures of each of these constructs were utilized to derive latent variables that were composed of only the shared variance among the three specific measures for each factor (i.e., Inhibition, WM, and CF). This use of a latent-variable procedure marked a significant improvement over many existing studies as many only utilized a single measure of EF leaving the study susceptible to skepticism as to whether or not the researchers actually captured the construct of interest.

Miyake and colleagues (2000) make the argument that use of latent variables, as opposed to individual, manifest variables of each EF component, is a promising approach for untangling the task-impurity problem that may arise from utilizing only a single measure of a given construct. The task-impurity problem is characterized by the contamination of other irrelevant skills/abilities within measures specifically designed to capture only a particular component of EF. For example, tasks that are purporting to measure WM may be contaminated by demands on Inhibition, making it difficult to determine what the task is actually measuring. Unfortunately, many studies investigating EF are undermined by task-impurity issues which leave researchers questioning the validity of various single measures of EF.

The results of the Miyake et al. (2000) CFA revealed that after analyzing Inhibition, WM and CF as three distinct factors, a full three-factor model provided the best fit for the data relative to both a two-factor model and a one-factor model of EF. These results seemingly answered the debate in favor of a diversity theory of EF. In other words, EF appeared to be best represented by the model that included all three separate components of EF, as opposed to any other nested combination. Despite this finding, the authors stated that although the three factors were separable at the latent-variable level, they shared some underlying commonality, as evidenced by moderate correlations among the different components (i.e., Inhibition, WM, and CF). This left the researchers to conclude that the evidence supported a *Unity/Diversity Framework* of EF; in that the unity is demonstrated by the three factors/components sharing a certain degree of variance that represents an overall EF, while also demonstrating diversity in that the factors are only moderately correlated among each other as the absence of a perfect correlation indicates the factors are measuring something different.

Continuing support for the *Unity/Diversity Framework* of EF has come from work with numerous populations including children, adolescents, older adults, and clinical populations (Friedman et al., 2008). For example, using a sample of 582 adolescents (aged 16 – 17 years) from 293 same-sex twin pairs, Friedman and colleagues (2008) were able to replicate the same three factor structure evident in the original Miyake et al. (2000) study. Again, this three-factor model provided a significantly better fit relative to both a two-factor and one-factor model of EF.

Subsequently, research conducted with preadolescent preterms (i.e., preterm infants that have aged into late childhood) has also demonstrated the *Unity/Diversity Framework's* efficacy. Specifically, Rose, Feldman, and Jankowski (2011) investigated whether the three components of EF identified by Miyake and colleagues (2000) were distinguishable among 11-year-old preadolescent preterms as well as normal controls. The researchers found that the three-factor model of EF (Miyake et al., 2000) provided an excellent fit to the data for both groups of preadolescents, with the three-factor model fitting better than all nested alternatives (i.e., combining two latent factors into one or removing a factor). Overall, research testing the utility of the *Unity/Diversity Framework* across numerous age groups has consistently supported the three-factor model.

An Emerging Conceptualization of Executive Functioning

Since publication of their frequently cited paper in 2000, Miyake and Friedman (2012) have continued to investigate the *Unity/Diversity Framework* and how the different factors (i.e., Inhibition, WM, and CF) interact with one another to contribute to the overarching factor, Common EF. Using other studies that have investigated this same three-factor model (Friedman et al., 2008, 2011), the authors suggested that after breaking down the explained variance that each of the three separate factors contributes to Common EF, there is no unique variance left for

Inhibition to explain. In fact, the inhibition factor appears to have an almost perfect correlation with the Common EF factor.

Research conducted with monozygotic and dizygotic twins has supported this unusually high correlation between Inhibition and overall EF, finding that the Inhibition latent variable had a loading of 1.0 on the Common EF factor (Friedman et al., 2008). The authors made note that their finding does not mean that Inhibition does not exist as a separate ability, but instead Inhibition's variance is completely explained by the shared variance of the other two factors. Research conducted with toddlers (Friedman, Miyake, Robinson, & Hewitt, 2011) has also demonstrated an exceptionally high or perfect correlation between Inhibition and overall EF, as all of the variance of the Inhibition-latent factor was completed explained by Common EF. What these findings demonstrate is that Inhibition appears to be accounted for by what is shared across WM and CF.

The role of Inhibition. Overall, Inhibition seems to be important to EF generally, across each of the interrelated constructs that contribute to EF (i.e., WM and CF). In other words, Inhibition may be embedded within, and a required component of all other EFs such as WM and CF. This might explain why the variance that is common among the latent variables of WM and CF seemingly encapsulates the variance that Inhibition captures. To be clear, the unusually high correlation between Inhibition and Common EF that has been demonstrated by numerous studies (e.g., Friedman et al., 2008; Miyake & Friedman, 2012) appears to suggest that one's success with Inhibition may contribute to both, WM and CF. The extent to which this is true is currently unknown.

Overall, more research is necessary to parse out the complex role each component makes in the unfolding of overall EF. For now, the three-factor model (i.e., Inhibition, WM, and CF) originally proposed by Miyake and colleagues (2000) appears to have the strongest empirical support as evidenced by numerous factor analysis studies that have demonstrated the three-factor model to have the best fit across numerous populations. Regardless of the nature of Common EF, it is clear that Inhibition, WM, and CF contribute to it in similar and unique ways.

The Relationship between Executive Functioning and other Domains of Functioning

There is a vast research base supporting the importance of EF on numerous domains of functioning. Researchers have examined how EF and its individual factors contribute to academic achievement, with specific emphasis on the areas of math and reading (Morgan et al., 2017). Other areas of study have included the relationship between EF and various aspects of social-emotional functioning such as the presence of externalizing and internalizing issues (Sulik et al., 2015). The following presents an in-depth review of the literature within each of these domains.

Relationship between EF and academic achievement. In their large, nationally-representative study, Best et al., (2011) examined the influence of EF on academic achievement over a wide age range (5 to 17 years) of students. The results indicated that regardless of age or subject (i.e., math vs. reading), EF showed a stable, moderate to moderately-large relationship with academic achievement. A more recent longitudinal study using a nationally-representative sample of first-grade children found a similar relationship, with the authors stating that EF deficits in kindergarten predicted both reading and math difficulties in first grade even after controlling for potential confounds such as children's prior history of reading or math difficulties, prior behavioral self-regulation skills, and family SES (Morgan et al., 2017).

A recent meta-analysis conducted by Yeniad, Malda, Mesman, van Ijzendoorn, and Pieper (2013) has further confirmed the importance of a specific component of EF, CF, finding a significant and small effect size (*r* = .26) between cognitive shifting and performance on math tests, as well as a significant and small effect (*r* = .21) between cognitive shifting and reading performance regardless of children's SES, gender, age, or grade level. It is important to note that overall intelligence (i.e., IQ) was found to be a significantly stronger contributor to academic performance than CF; however, IQ and CF were also significantly related to each other. This latter finding suggests that CF may be important to, or a byproduct of, overall intelligence.

Relationship between EF and social-emotional functioning. Similar to the plethora of research examining the relationship between EF and academic achievement, EF and social-emotional functioning has been extensively studied. For example, a higher level of EF has been shown to manifest in protective, adaptive behaviors for children. More specifically, relative to children with low CF, children with high CF have been shown to exhibit more cooperative social behaviors (e.g., offering and accepting help on a puzzle, letting another child attach a puzzle piece to the section a child is working on) and less non-cooperative behaviors (e.g., stealing a puzzle piece from the hands of the partner, denying a puzzle piece to the other child) during a play situation where two children were given a puzzle to share (Ciairano et al., 2006). Further, CF has been shown to predict social understanding in 7- to 12-year-olds after controlling for various confounds such as age, vocabulary, WM, and Inhibition (Bock et al., 2015). For example, higher CF in this study predicted a child's competence with answering comprehension questions about several social vignettes that depicted characters who were pretending, lying, joking, using figure of speech, and bluffing.

In addition, research conducted with preschoolers has established the importance of EF with predicting problem behaviors, finding that EF measured at age three significantly predicted aggregate problem behavior scores at four years old (Hughes & Ensor, 2008). Research has also found links between EF and various social-emotional concerns such as the presence of antisocial behavior, internalizing disorders, and externalizing disorders (Hughes & Ensor, 2011; Ogilvie et al., 2011).

Executive function and internalizing disorders. Internalizing disorders are generally characterized by symptoms of anxiety and depression, with the child presenting with higher levels of negative affectivity than what is considered typical (Kovacs & Devlin, 1998). Similar to the literature investigating the link between EF and externalizing disorders, the link between EF and internalizing disorders has strong support as well. In their longitudinal study using a latent-variable approach, Hughes and Ensor (2011) found that developmental gains in EF (i.e., growth), as opposed to the degree of EF exhibited at a single point in time, significantly predicted not only EBP, but also internalizing problems from four to six years of age. More specifically, children who exhibited less development in EF were more likely to have emotional and behavioral issues. Interestingly, high gains in EF were also found to have a positive effect on self-perceived academic competence illustrating a protective component that higher EF may offer.

Another study illustrated the importance of EF and its role in the presence of internalizing symptoms (Tan, 2011). Utilizing a Philippine student sample, the researcher derived composite variables for Inhibition, WM, and CF in efforts to see how they differentially related to both anxiety and depression. In particular, the composite variables were derived from three distinct sources. The first source was a performance-based measure directly administered to the child. In other words, the child completed: The Color-Word Interference subtest from the Delis-Kaplan Test of Executive Function System (DKEFS; Delis, Kaplan, & Kramer, 2001) as a measure of Inhibition, the Working Memory Clinical Cluster (i.e., the Numbers Reversed and Auditory

Working tests) of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III-COG) as a measure of WM, and the Wisconsin Card Sorting Test (WCST) as a measure of CF. The second source was a subscale (Inhibit, Shift, and Working Memory scales) from the Behavior Rating Inventory of Executive Functions-Self-Report Version (BRIEF-SR; Gioia, Isquith, Guy, & Kenworthy, 2000). Finally, a subscale (Inhibit, Shift, and Working Memory) from the Behavior Rating Inventory of Executive Functions-Teacher Form (BRIEF-TR) was completed by a classroom advisor/teacher.

Results from this study indicated that both WM and CF had significant correlations with anxiety and depression. Interestingly, a significant correlation was not found between Inhibition and either anxiety or depression. Despite this, Inhibition was significantly correlated to both WM and CF. This suggests that Inhibition's influence on symptoms of anxiety and depression is mediated by both WM and CF. Regarding the direction of the relationships, negative correlations between WM, CF, and both anxiety and depression were found suggesting that as internalizing problems increased, EF decreased. Despite this rationale, this study was correlational in design and there was no clear temporal ordering to the variables. In other words, the reverse direction of causation is just as plausible, with low EF leading to higher symptoms of anxiety and depression. Finally, being that the sample was from the Philippines, generalizability of this study is quite limited given the lack of representation from different cultures and ethnicities within the sample.

Executive function and EBP. Definitions from the literature concerning EBP all converge on behaviors characterized by aggression, impulsivity, defiance, disruptiveness, and overactivity (Hinshaw, 1992). Common ways of measuring the presence of EBP in children include the use of rating scales (e.g., Child Behavior Checklist [CBCL], Achenbach & Rescorla, 2001; Behavior Assessment System for Children, Third Ed.[BASC-3], Reynolds & Kamphaus, 2015; Social Skills Rating System [SSRS], Gresham & Elliott, 1990) and structured observation systems, the degree of juvenile delinquency, and the use of nosological systems such as the Diagnostic and Statistical Manual of Mental Disorders – Fifth Ed. (DSM-5; American Psychiatric Association, 2013).

The presence of EBP during childhood has long been found to lead to numerous maladaptive outcomes later in life such as, academic underachievement, antisocial behavior, peer problems, and substance abuse (Hinshaw, 1992). Common diagnoses that fall within this category include: ADHD, CD, and ODD. Per the DSM-5 (American Psychiatric Association, 2013), current prevalence estimates for each of these disorders are estimated at 5% of all children, 3.3% of all children, and 4% of all children, respectively. One common feature of all these disorders is the presence of emotional impulsivity (EI) and deficient emotional self-regulation (DESR; Barkley, 2014). EI is defined as the speed of which an individual reacts to a stimulus with a primary emotion, often characterized as negative in nature (e.g., anger). DESR, on the other hand, is defined as an individual's inability to inhibit strong positive or negative emotions, engage in self-soothing behaviors that reduce physiological arousal, divert attention, and organize their behaviors towards a more adaptive goal (Gottman & Katz, 1989 as cited in Barkley, 2014). Given these problems with managing one's own behaviors, any investigation into the etiology of EBP must include EF as its definition is directly connected to regulating maladaptive emotions and behaviors.

Support for the link between EF deficits and EBP has come primarily from research utilizing children and adolescents with ADHD. A meta-analysis investigating EF deficits among preschool children (i.e., three through six years) with EBP established that EF plays an early role in the presence of problem behaviors during the preschool years (Schoemaker et al., 2013). Using the Miyake et al. (2000) framework, the researchers found medium effect sizes for overall EF (r = 0.22) and Inhibition (r = 0.24), and small effect sizes for WM (r = 0.17) and CF (r = 0.13). These results indicated that preschoolers with low EF generally showed higher levels of acting-out behavioral problems.

Research conducted with clinical populations has also provided valuable insights. The consensus appears to be that EF deficits are ubiquitous within cases of ADHD regardless of age, with significant negative outcomes stemming from these EF deficits (Antshel, Hier, & Barkley, 2014; Barkley, 2014; Barkley & Fischer, 2011). For example, a study conducted by Barkley and Fischer (2011) found that ratings of EF were significantly related to ADHD severity among adults, with more EF deficits being associated with more symptoms of the disorder. Further, children and adolescents with ADHD have been found to have significantly more EF deficits relative to healthy controls (Biederman et al., 2004). Overall, the current literature base points to a pivotal role for EF in the presence of EBP across the lifespan.

The Relationship between Executive Functioning and Demographic Variables

Individuals can vary widely in the degree to which they exhibit EF. As such, understanding what contributes to the varying levels of EF is crucial in order to aid those who are exhibiting EF deficits or developmental delays. Moreover, given the wide number of human behaviors that EF is thought to influence (e.g., planning, set shifting, working memory, self-regulation, etc.), it is even more important to understand what the key variables are that contribute to the development and manifestation of EF as delays or deficits in these domains can have far reaching negative consequences. Of the innumerous potential key predictor variables, the EF literature offers robust evidence as to the degree to which certain demographic variables (e.g., race, sex, and SES) contribute to the unfolding of EF.

Race and executive functioning. There is consensus among researchers that racial inequalities are present when examining group differences in cognitive abilities; however, determining the causal mechanisms (i.e., gene vs. environment) for these observed differences has been a controversial debate (Cottrell, Newman, & Roisman, 2015; Roth, Bevier, Bobko, Switzer III, & Tyler, 2001). Some of the potential causes that have been discussed in the literature include parenting factors, a disproportionate distribution of racial minorities in low SES neighborhoods, and access to learning materials in the home (Cottrell et al., 2015). When trying to observe racial differences in EF, the same debate (i.e., genes vs. environment) persists when attempting to discern whether any group differences that are observed are in fact genuine differences between people from different racial backgrounds or simply a byproduct of societal inequality.

In one of the only studies examining racial differences in EF among children, Little (2017) provided similar findings to the years of research investigating racial differences in cognitive abilities, confirming that disparities exist between children from differing racial backgrounds. Using the same nationally-representative sample that the current study will employ, Little (2017) investigated whether race and SES predicted Academic Achievement, WM, and CF from kindergarten to second grade. The results indicated that beginning in Kindergarten, students who were Black or Hispanic were found to exhibit significantly lower Academic Achievement, WM, and CF when compared to White peers. For example, Hispanic students were found to perform 0.59 standard deviations lower on the WM task (i.e., Numbers Reversed), on average, than the white comparison group.

In a more positive light, the gaps in WM and CF that were present in Kindergarten were found to narrow over time at a faster rate than the gaps that were observed in academic achievement (Little, 2017). For example, Little reported that the gap in WM present at Kindergarten more than halved by second grade to 0.27 standard deviations as opposed to the gap in Math shifting from 0.64 standard deviations in kindergarten to 0.54 standard deviations in second grade. Nonetheless, although the gaps in EF more than halved by second grade, they still persisted at a statically significant level, meaning that the students who identified as Black or Hispanic still performed worse than students who identified as White, albeit to a lesser degree.

Sex and executive functioning. ADHD has been characterized as a disorder of executive dysfunction due to research demonstrating similarities between those with ADHD and those suffering frontal lobe damage (Barkley, 2014; Seidman et al., 2005). In particular, those with ADHD often exhibit impulsivity, hyperactivity, and distractibility. When examining sex differences in the prevalence of ADHD, males are disproportionately represented when compared to females, with one study suggesting a male bias as high as 10:1 in clinic-referred samples (Biederman et al., 2002).

Although males appear to exhibit a disorder of executive dysfunction (i.e. ADHD) at a rate far higher than females, research examining sex differences in EF has found little evidence supporting a significant difference between sexes in the degree to which they exhibit EF. One study in particular examined whether 9-17-year-old girls and boys, with and without ADHD, differed significantly in EF (Seidman et al., 2005). After controlling for age, SES, the presence of a learning disability, and psychiatric comorbidity, the results indicated that girls and boys with ADHD differed significantly from healthy controls; however, they did not differ from each other in terms of EF. Additionally, healthy girls and boys did not differ significantly from each other in regard to EF.

Further support for the lack of significant sex differences in EF comes from a more recent study conducted on a sample of nine through 12-year-old school children (van Tetering & Jolles, 2017). Instead of relying on individual testing of EF, researchers administered an observer-rated questionnaire (i.e., both teacher and parent completed the questionnaire on the child) and examined age-related changes in EF and whether sex contributed to the changes in EF that were observed. Although significant growth in EF was observed among the children included in the study, no significant sex difference was found in the total EF score that was derived from the teacher and parent ratings.

Further evidence supporting the lack of significant sex differences in EF comes from research utilizing the Dimensional Change Card Sort (DCCS), the measure of CF being employed in the current study, in young children. In one study (Diamond, Carlson, & Beck, 2005), 57 preschool children (27 males, 30 females) were administered the standard DCCS at three ages (2.5, 3, & 3.5 years). No significant sex differences were found at any of the tested ages. Another study (Hongwanishkul, Happaney, Lee, & Zelazo, 2005) conducted utilizing a sample of 98 children between the ages of 3.0 and 5.9 years yielded similar results, finding no significant main effect of sex and no significant age x sex interaction. Taken together, the evidence demonstrates that males and females do not significantly differ in CF during the childhood years.

Despite the lack of evidence supporting significant sex differences in CF throughout childhood, research investigating sex differences in WM are well documented. One particular distinction that is important is that between visual-spatial WM and verbal WM. A recent meta-analysis conducted by Voyer, Voyer, & Saint Aubin (2016) utilizing 98 samples with participants ranging from 3 years to 85 years revealed a small, significant male advantage beginning in the 13-17-year-old group on measures of visual-spatial WM. This finding points to the small advantage that has been well-documented for males in their ability to handle visual-spatial information.

In contrary to the well-established male advantage for visual-spatial WM tasks, evidence of a sex difference in verbal WM is mixed. For example, a recent study investigated the degree to which WM, in general, contributes to the well-established sex differences in mental rotation and spatial-visualization ability (Kaufman, 2007). Spatial ability, spatial WM, and verbal WM were all assessed via individual testing. Results supported the conclusion that males outperform females on spatial ability and spatial WM tasks; however, no significant difference was found between males and females on the verbal WM test.

In another study investigating both visual-spatial and verbal WM tasks among 274 healthy college students, researchers found significant sex effects and qualitative differences in brain areas that were recruited for various WM tasks; however, mixed evidence regarding verbal WM was found yet again (Zilles et al., 2016). In particular, males were found to outperform females on visual spatial WM and verbal WM with high memory load. Despite the significant difference being found in the verbal WM with high memory load task, no significant sex differences were found on the verbal WM with low memory load task.

Finally, in another study (Robert & Savoie, 2006) investigating sex differences in verbal and visuospatial WM among 19-25-year-old men (n = 50) and women (n = 50), researchers found convergent support for the widely-accepted finding that males outperform females on visuospatial WM tasks; however, null findings between the sexes within the digit-span task, the same type of task the current study will employ, provides further evidence suggesting no significant difference between the sexes in verbal WM ability. In particular, males and females performed equally well when asked to recite auditorily-presented words in both forward and backwards directions. In conclusion, on tasks of verbal WM, the findings are mixed as to whether sex differences exist but data seems to suggest that no significant sex difference is present.

Socioeconomic status and executive functioning. It is well known that early exposure to environmental adversity is predictive of a host of cognitive and physiological difficulties later in life (Letourneau, Duffett-Leger, Levac, Watson, & Young-Morris, 2013; Raver, Blair, & Willoughby, 2013). In order to objectively measure an individual's early exposure to adversity, or lack thereof, researchers have often relied upon SES. SES is a composite variable typically derived from parental income, parental education, and parental occupational prestige (Ursache, Noble, & Blair, 2015). Consequently, SES serves as an indicator of the environment parents are able to provide for their child.

Numerous studies have found that children from lower SES families perform significantly worse on measures of EF (Raver et al., 2013; Sarsour et al., 2011; Ursache et al., 2015). For example, using a community sample of 60 ethnically and socioeconomically-diverse, English-speaking families, Sarsour et al. (2011) found that among the 8-12-year-old children tested, those from lower SES families performed significantly worse on measures of Inhibition, WM, and CF. In another study using a larger sample size of 1,292 children coming from predominantly low-income families, Raver et al. (2013) demonstrated that financial hardship uniquely predicted performance on an aggregate measure of EF that included tasks that tapped into Inhibition, WM, and Attention Set Shifting (i.e., CF). In particular, the Raver et al. (2013)

study indicated that exposure to poverty early in life and exposure to chronic poverty strongly predicted worse performance on measures of EF with the degree and frequency of exposure to poverty and to poverty-related hazards from infancy to early childhood explaining 22% of the variance in EF as early as 2 years old. Finally, in a more recent study investigating the degree to which SES influences a child's EF, Ursache et al. (2015) demonstrated a positive correlation between SES and EF using a socioeconomically diverse sample of children aged 6-12 years. In particular, these researchers found a strong positive correlation between SES and EF ($r = .31$; $p < .01$), with increases in SES being associated with increases in EF. In conclusion, the data unequivocally shows that socioeconomic disparities exist when measuring EF in childhood.

Developmental Trajectory of Executive Functioning's Components

Much of the literature on the development of EF has focused on early childhood (i.e., three to six years), providing a rigorous knowledge base when trying to identify trends within the various components of EF during this time frame (Best & Miller, 2010). Despite the abundance of research for this age group, the EF literature contains many disparate views related to the definition of EF, the structure and organization of EF, and how to measure EF and its components (Best & Miller, 2010). Given the increasing attention paid to early childhood, these problems are amplified when trying to examine EF during later time frames such as middle childhood (i.e., seven years to ten years) and adolescence, as these time frames have not been as extensively studied. This leads to difficulties in summarizing research and making clinical judgements as to how EF manifests at different ages, especially during middle childhood and beyond. Despite these difficulties, researchers (Best & Miller, 2010; Best et al., 2009) have recently attempted to simplify the complications surrounding the developmental nature of EF through reviews of the literature utilizing a common framework that has empirical support. The following sections in this document continue this tradition utilizing the *Unity/Diversity Framework* that was previously introduced (Miyake et al., 2000) so to keep a consistent conceptual framework throughout and add to current trends within the literature base.

Working Memory. According to Diamond (2006), WM reveals itself as early as infancy, as infants aged eight to 12 months are able to remember the hidden location of desired objects and elect not to use previously unsuccessful search methods for finding hidden objects. Development of WM then continues through the preschool years (i.e., three to six years), with improvements on numerous behavioral measures of WM being found among children as young as four years old (Gathercole, Pickering, Ambridge, & Wearing, 2004). Development of WM through adolescence has also been clearly noted as numerous studies have found a linear trend of development through adolescence (Best & Miller, 2010). For example, Gathercole and colleagues (2004) have found continuing improvements with WM all the way through age 14 years as measured by numerous tests of WM (i.e., Backward Digit Recall, Word List Recall, Nonword List Recall, Block Recall, and Visual Patterns Test). Additionally, a longitudinal study noted behavioral improvements in various aspects of WM (e.g., recall-guided action, visual-spatial recall, and strategic self-organization) from nine years to 17 years (Luciana, Conklin, Hooper, & Yarger, 2005). In general, the development of WM is characterized by long term improvements through adolescence, where performance appears to reach adult levels (Best & Miller, 2010).

Neuroscience research has provided corroborating evidence to the behavioral development of WM that has been identified by changes within a key brain region associated with WM, the prefrontal cortex (PFC). In their review, Best and Miller (2010) found numerous studies that utilized functional magnetic resonance imaging (fMRI) to indicate quantitative and qualitative changes in the PFC from childhood through early adulthood. For example, Scherf, Sweeney, and Luna (2006) found changes between childhood and adolescence in both the location of neural activation within the frontal lobe, and the amount of activation as related to performance on WM tasks. Specifically, premotor regions of the brain were relied upon more heavily during childhood with increasing reliance on more frontal regions, such as the dorsolateral prefrontal cortex (DLPFC), occurring during adolescence. The shift to the utilization of the frontal regions of the brain indicates that as people age and utilize WM more effectively, they begin to recruit increasing amounts of activity from regions that have been previously identified as WM circuits (e.g., PFC and DLPFC; Scherf et al., 2006). Another study conducted by Kwon, Reiss, and Menon (2002) found increased activity in regions of the PFC between the ages of seven years and 22 years while completing visuospatial WM tasks further supporting the shift to more frontal regions as people age.

Interestingly, other researchers (e.g., Conklin, Luciana, Hooper, & Yarger, 2007) have hypothesized domain-specific or processing-specific differences in the rate of development of WM. Domain-specific differences entail a differential developmental trend between WM tasks characterized as either entirely verbal (i.e., tasks requiring the processing of language) or visuospatial (i.e., tasks that require visual recollection of where objects are in space) in nature. In contrast, processing-specific differences are characterized by the degree of processing demands relative to the task, with the hypothesis being that tasks requiring higher amounts of processing (e.g., tasks requiring strategic self-organization) would develop later than tasks that require less processing (e.g., tasks requiring recall-guided action).

Using a sample of typically-developing 9 through 17-year-old students, Conklin et al. (2007) found a similar developmental trend for verbal WM tasks relative to visuospatial WM tasks. Thus, the evidence failed to support the domain-specific hypothesis of WM development. In order to test the processing-specific hypothesis, they completed a principal components factor analysis to determine the best fitting model of WM among the numerous WM measures they administered (e.g., span tasks, self-ordered search tasks, and recognition tasks). Results from the factor analysis revealed that the grouping of tasks was largely based on task demand (i.e., maintenance, manipulation, or self-organization) as opposed to being grouped by the domain the tasks measured (i.e., verbal or spatial). For example, all of the span tasks represented one factor whereas the self-ordered tasks represented another factor even though both groups of tasks contained measures primarily verbal or visuospatial in nature.

Further support for the processing-specific model of WM has come from research conducted by Luciana et al. (2005). In their study, the researchers hypothesized that the development of WM from nine years to 20 years would be characterized by a hierarchical progression with simpler tasks reaching adult-performance levels before more complex tasks. In particular, participants were administered a series of WM measures that varied in executive demand. At the lower end of the hierarchy of difficulty, the participants were asked to encode pictures of human faces and then select the face they saw among two choices after a five-second delay. Given this was a forced-choice test with only a recognition component, it was thought to exert low executive demand (Luciana et al., 2005). Next, the participants were administered a spatial-delayed response task where they had to pinpoint the location of a dot that was flashed on a computer screen either 0.5 seconds or eight seconds before their response was allowed. This task required the individual to maintain the location of the dot in their WM, and then recall the location of the dot using a touch pen.

After the spatial delayed response task was administered, a spatial memory span task was given to each participant. During this task, the participant observed a trained administrator point to a sequence of three-dimensional blocks on a flat board. The participant was then asked to reproduce the same sequence by touching the blocks. This task represented more difficulty than the spatial-delayed response task because the participant was required to maintain multiple units of information (i.e., the sequence of blocks) simultaneously as opposed to single units of information (i.e., the location of a single dot). The final task that was administered to participants hypothetically required the greatest executive control given that it was a self-ordered search test. In particular, participants were required to locate hidden tokens, with search complexity being varied by the number of locations (i.e., three to eight locations) the participant had to search. Forgetting errors would be recorded if a participant searched a location where he or she had previously already found a token. In addition, a strategy score was derived that indicated the degree to which the participant utilized a strategy for searching the locations. For example, a participant who followed a systematic pattern of searching unchecked locations would score high in strategy as opposed to someone following a simple trial-and-error approach.

Analysis of the data revealed strong support for the processing-specific hypothesis of WM development. In particular, their results indicated that recall-guided action for single units of information (i.e., spatial delayed response task) develops until 11 to 12 years, whereas recall for multiple units of information does not reach peak levels until approximately 13 to 15 years. Finally, development of strategic self-organization (i.e., self-ordered search test) was shown to develop until ages 16 to 17 years, with performance not differing significantly from the 18 to 20-year-old group. This illustrates that development of WM follows differential paths depending on the degree of activation the task requires.

Overall, the findings from the Conklin et al. (2007) study indicated no domain-specific differences in the development of WM; however, a factor analysis of the data revealed a process-specific structure of WM leading the team to conclude that the content to be processed does not matter so much as the degree of processing when trying to understand the developmental path of WM. Furthermore, it appears that a child's WM for visuospatial and verbal tasks appears to develop simultaneously; however, the child's ability to handle increasingly complex WM tasks develops hierarchically and through adolescence. This latter point was demonstrated by the Luciana et al. (2005) study which showed a clear hierarchical development of WM that lasted until 17 years, where participants were able to handle high executive demand at adult-performance levels.

Cognitive Flexibility. Similarly to the development of WM, the literature surrounding the development of CF points to a protracted development that shows improvements beginning in early childhood (i.e., three years to five years) and continuing into adolescence (i.e., 15 years) where CF begins to reach adult-levels (Cepeda, Kramer, & Gonzalez de Sather, 2001; Davidson, Amso, Anderson, & Diamond, 2006; Huizinga, Dolan, and van der Molen, 2006). In their review, Cragg and Chevalier (2012) found that most three-year-old children fail to successfully switch to an alternate sorting dimension as measured by a well-studied CF measure called the DCCS. On this task, participants are required to sort cards according to a specific dimension (i.e., color). After establishing this initial sorting dimension, the participant is then required to shift to an alternate sorting dimension (i.e., shape). The number of cards sorted correctly is then scored and the difference between both phases indicates shift cost (i.e., increase in the number of errors during the second phase when the participant was required to sort by a different rule).

In contrast to the three-year-old's difficulties switching on the DCCS, four and five-year-old children are often more successful when asked to sort by the alternate rule set (Best & Miller, 2010; Cragg & Chevalier, 2012). The consensus seems to be that the three-year-old's lack of CF reflects two phenomena. First, they have difficulty disengaging from the previous mental ruleset that was required to initially sort the items leading them to have perseverative errors on the second phase characterized by misusing that initial rule for the alternate sorting. Second, the three-year-olds also have difficulty activating a previously ignored stimulus to sort by an alternate rule set (Cragg & Chevalier, 2012). In other words, shape was an irrelevant stimulus during the initial phase; however, shape becomes relevant during the second phase. In general, these patterns indicate that the successful manifestation of CF can be elusive for younger children under the age of 4.

But what explains why four- and five-year-old children can handle flexibly switching to a second sorting dimension while three-year olds cannot? In their review, Cragg and Chevalier (2012) highlight the importance of goal setting for successfully exhibiting CF. Goal setting is thought to be highly dependent on verbal ability; more specifically, the areas of inner speech and verbal memory. Research conducted by Chevalier and Blaye (2009) illustrated that the effects of cue transparency (i.e., providing auditory or visual cues that signal the sorting dimension that should be used on each trial) on improving CF performance diminish as the participants' ages increase. For example, providing auditory or visual cues helped increase the performance of children aged five through nine years; however, the positive effects of providing cues decreased over time. They hypothesized that this resulted because older, more developed participants can rely on more efficient inner speech strategies rather than relying solely on external cues. Thus, the older children relied on the cues less because they had adequate sub-vocal rehearsing processes to handle the task demands of the DCCS. Although no research has been published examining this connection directly, it seems likely that an increased efficiency with verbal ability would aid performance on CF tasks, given the need to set goals dependent on the rules of the task.

One important distinction in the measurement of CF concerns the age of the population that is being studied. As already stated, the DCCS is a widely used measure for early childhood with CF being measured by the number of items that are sorted correctly. In contrast, older children and adults are administered task-switching paradigms given that performance on the standard DCCS ceilings out around five years of age (Best & Miller, 2011; Cragg & Chevalier, 2012; Davidson et al., 2006). With task-switching paradigms, accuracy and reaction time are often recorded while the participant switches between different response rules that are dependent on various cues within a presented stimulus. This change in the dependent measure of CF often reveals further intricacies within older participants. In particular, researchers have noted that adults will slow their response times (RT) down in order to be more accurate as opposed to children who will not (e.g., three years through five years; Cragg & Chevalier, 2012; Davidson et al., 2006). For example, Davidson et al. (2006) demonstrated this RT-accuracy tradeoff using a sample of 325 participants aged four through 45 years. Specifically, adults (mean age = 26.30 years) were significantly more accurate than children and adolescents on computerized tasks that measured switch cost; with adults demonstrating a significantly smaller switch cost relative to children and adolescents aged six through 13 years. This indicates that even the 13-year-old participants had not reached adult levels of CF as measured by the difference in accuracy during the post-switch phase. Further, the adults' significantly higher accuracy was accompanied by slower RTs on more difficult trials whereas the four-year-old children showed a more stable RT indicating impulsivity, as they failed to slow down on trials that were more difficult.

Relative to the development of Inhibition and WM, researchers have observed that children often display less improvements in CF during early childhood (Best & Miller, 2010). One theory concerning this delayed development suggests the incorporation of WM and Inhibition as building blocks for CF. More specifically, successful CF may require WM and Inhibition concurrently due to the complex nature of the tasks that are used to assess CF. As Best and Miller (2010) pointed out in their review, most tasks measuring CF require the participant to hold a response rule set in WM, inhibit that first rule set, and then formulate the new rule set based on feedback from the task. Thus, it appears that CF must necessarily develop slower than WM and Inhibition as it is built upon these two abilities. Despite this, few studies have tested this hypothetical, hierarchical development of CF empirically.

Inhibition. Inhibition has been a source of controversy due to disagreement among researchers about which tasks measure Inhibition and which measure CF (e.g., Wisconsin Card Sort vs. DCCS; Best & Miller, 2010). Even measures that are touted as Inhibition tasks suffer from the task-impurity problem where other skills/abilities are present in the task making it impossible to decipher whether growth is occurring in these irrelevant skills/abilities or Inhibition. The lack of consensus concerning how to measure Inhibition seems to hit upon Miyake and Friedman’s (2012) developing framework that states Inhibition may not be a separate factor of Common EF due to the near perfect correlation it has with the latent variable of Common EF. The difficulty in measuring Inhibition seems to provide supporting evidence that Inhibition is crucial among all aspects of EF and therefore would be difficult to parse out among various tasks that purport to measure EF. Despite the ambiguity surrounding how to measure Inhibition, the developmental research does provide some insights.

According to their review of the literature, Best and Miller (2010) found numerous studies depicting different ages of mastery on Inhibition tasks depending on the distinction between simple-response inhibition and complex-response inhibition. Simple-response inhibition is considered more of a pure form of Inhibition as the demand on WM is reduced, requiring only the inhibition of a prepotent response (e.g., a child delaying eating an edible reinforcer). Success is determined by simply not exhibiting the dominant response that is elicited. In contrast, complex-response inhibition not only requires the inhibition of a dominant response, but also the production of an alternative response in exchange; this simultaneously forces the individual to hold a goal in present awareness in order to guide the selection of an alternate response. Based on this definition of complex inhibition; however, the distinction is blurred with CF, which also requires the inhibition of a dominant response in conjunction with the production of an alternate response. Despite this ambiguity, it can be hypothesized that complex-inhibition tasks would lead to later ages of mastery since both successful Inhibition and CF are hypothetically required for these tasks.

Tasks within the complex-inhibition category include the Day-Night Task and Luria's Hand Game (Best & Miller, 2010). The Day-Night Task requires the child to inhibit a dominant verbal response in favor of an alternate verbal response when presented with visual stimuli (i.e., saying "night" in response to a picture of the sun). Luria's Hand Game requires a child to make hand gestures (e.g., make a fist) in response to visual stimuli (e.g., picture of fingers). Both of these tasks have been described by researchers as conflict tasks, meaning that the child is required to produce a conflicting response when presented with visual stimuli. For example, the correct answer is not the intuitive answer but one that contradicts the stimuli that is presented.

Research using primarily complex-inhibition tasks has generally shown that significant development happens during the preschool years (Hughes, 1998). With tasks that require a verbal response of some sort (i.e., Day-Night Task), improvements have been noted into middle childhood (Best & Miller, 2010), likely due to the increase in the verbal skills of the child being tested. In other words, the improvements that have been noted within Inhibition tasks during middle childhood are difficult to substantiate, as the measures that are typically used suffer from the task-impurity problem; this means that researchers cannot be certain that observed improvements are solely due to increases in Inhibition as opposed to simple increases in other skills/abilities that would improve performance on these tests.

After five years of age, the research findings regarding the development of Inhibition are mixed, with some finding minimal to no significant improvement, and others finding improvement after eight years of age (Best & Miller, 2010). It is not clear why the findings are mixed in later childhood, but one explanation could be the different task demands of the measures that are used in later childhood. As already stated, it is hard to say whether or not the improvements that are noted in Inhibition during later childhood are simply due to improvements in other areas such as verbal ability or computer skills (since a lot of tasks are computer-based). Additionally, the lack of consensus on how to measure Inhibition makes it hard to generalize findings across studies. Despite these difficulties, consensus appears to demonstrate that children show marked improvements between the ages of three years and four years on both simple- and complex-inhibition tasks. These improvements continue from age five years to eight years, especially on tasks that increase WM demand, as well.

Given the lack of consensus on how to measure Inhibition independent of other skills/abilities, and the lack of a pure Inhibition measure within the ECLS-K dataset, the current study did not include Inhibition within its model of EF. Consequently, the current model was unable to illuminate the specific influence that Inhibition has on WM, CF, or EBP. Future research will hopefully be able to disentangle the role that Inhibition plays in EF and the development of EF's components.

Current Study

The current study aims to add to the literature base concerning the link between EF and externalizing behavioral problems (EBP) during the elementary school years, using a large nationally-representative sample. In particular, the influence of WM and CF on EBP was investigated using the Early Childhood Longitudinal Study: 2010-2011 Kindergarten cohort (ECLS-K; Tourangeau et al., 2017). The rationale behind including WM and CF as measures of EF is supported by factor analysis studies that have consistently yielded WM and CF as two out of the three major components of EF (i.e., Inhibition, WM, and CF) among numerous samples (Friedman et al., 2008; Friedman et al., 2011; Miyake et al., 2000; Miyake & Friedman, 2012). Inhibition has not been included within the current study due to no Inhibition measure being present within the dataset as well as recent questions as to the role Inhibition plays in EF. In sum, the current study furthers the understanding of how specific components of EF (i.e., WM and CF) contribute to EBP during the elementary school years, as well as how WM and CF develop across time.

The goals of this study were to help inform school-based practitioners and researchers alike, helping to advance not only current assessment and intervention practices, but also our understanding of how EF unfolds throughout childhood. For researchers, the current project advances theory concerning how WM and CF interact developmentally as well as in relation to EBP. In particular, this is one of the first studies conducted that quantifies the degree to which WM influences the development of CF as well as each factor's relationship with EBP. Further, the developmental progression for both WM and CF is demonstrated from kindergarten through second grade, providing either convergent or discriminant evidence for studies that have already investigated the developmental timeline of these components of EF. For practitioners, the current study provides evidence of the relationship between EF and EBP during kindergarten through second grade.

The main hypotheses of the current study are:

(1) WM in the fall of kindergarten will have a significant negative relationship with EBP in the spring of second grade;

(2) CF in the fall of kindergarten will have a significant negative relationship with EBP in the spring of second grade;

(3) CF in the spring of kindergarten will mediate, to a significant degree, the relationship between WM in the fall of kindergarten and EBP in the spring of second grade;

Please refer to Figure 3 for a visual depiction of the main hypotheses in path format. In addition to the three main hypotheses presented above, two exploratory hypotheses will also be investigated:

(4) WM in the fall of kindergarten will significantly predict CF in the spring of kindergarten, as the development of WM is a hypothesized prerequisite for the development of CF;

(5) The relationship between each component of EF (i.e., WM and CF) and EBP will grow stronger as the temporal distance between the measures of EF and later EBP grow closer. In other words, WM and CF in the fall of second grade will have a stronger relationship

with EBP, as measure in the spring of second grade, than WM and CF in the fall of kindergarten.

One strength of the current study is its' ability to illuminate the contribution that WM makes to CF development (i.e., hypothesis four). WM has been postulated to be important to CF development by numerous researchers (Best & Miller, 2010; Blakey et al., 2016; Garon, Bryson, & Smith, 2008); however, no one has tested this using a nationally-representative sample of elementary-aged children. The current study adds to this knowledge base by using path analysis to determine if WM at earlier time points explains a significant portion of variance in later measures of CF (i.e., does WM contribute significantly to later CF). Moreover, indirect effects of WM through CF on teacher-rated EBP were also investigated. Understanding how these specific components of EF develop simultaneously, as well as how they relate to EBP longitudinally is necessary for future researchers and practitioners, alike.

Chapter III: Method

The research design and methodology are presented in the following four subsections: Database Overview, Variables, Data Preparation, and Analyses. In the first section, an overview of the existing database, Early Childhood Longitudinal Study Class of 2010-2011 (ECLS-K), is presented, as well as an explanation of the sampling design and data collection procedures. The second section identifies the key variables of interest included in the investigation and how the variables are operationally defined. In the third section, the data preparation phase is explained with special emphasis on analytical issues associated with complex survey data (e.g., weighting and missing data). Finally, the last section includes an explanation of the statistical techniques that were employed, as well as the path model that was tested.

Database Overview

This study utilized existing data from the ECLS-K. The ECLS-K follows a nationally representative sample of children from kindergarten through their elementary school years (Tourangeau et al., 2017). It includes multiple waves of data collection (i.e., fall and spring of each grade level from K [2010-2011] through fifth grade [2015-2016]), using multiple sources of data informants, such as parents, teachers, administrators, as well as direct assessments of children's functioning. In this study, direct child assessments and teacher ratings of social-emotional functioning were utilized to investigate the complex relationship between specific components of EF and the presence of EBP in the school setting, while controlling for familial background characteristics.

Participants. The ECLS-K consists of a nationally-representative sample of 18,174 kindergarteners from approximately 970 schools in the U.S. (Ansari & Purtell, 2017; Tourangeau et al., 2017). The sample was selected using a multi-stage, stratified sampling design that involved three steps in efforts to yield accurate estimates of national child-level data. First, the country was divided into 90 primary sampling units (PSUs) based on nearby geographic regions such as counties or groups of counties. The creation of PSUs was based on numerous criteria such as: proportional percentage of people who identify as Black and Hispanic, minimum number of five-year-old children was at least 380, the maximum distance between two points within a PSU was 100 miles, and the PSU formed within a state boundary. During the second stage of sampling, samples of public and private schools that educated children of kindergarten age (i.e., five years old) were selected from within each PSU. The decision to include specific schools was based on probabilities that took into account a desired oversampling of people who identify as Asian, Native Hawaiian, and other Pacific Islander (APIs). In other words, schools were selected if they had higher populations of these peoples in order to ensure adequate representation in the final sample. In the third stage, individual children were selected from within those schools based on an independent sampling strategy that broke individual school populations into two strata, one containing children identified as API, and another strata containing all other children. Children from the API strata were sampled 2.5 times more than children from the second strata (i.e., all other children) in efforts to meet the overall sampling goal of having a highly generalizable sample.

Variables

For the current study, all of the study variables were directly taken from the ECLS-K database. The control variables selected for this investigation include sex, race, and SES. The independent variables include WM and CF. Measures of WM and CF were obtained during the fall and spring of each year – kindergarten through second grade. Finally, the dependent variable is teacher ratings of EBP during the spring of second grade.

Control variables. Information concerning children's family background characteristics was collected during the fall 2010 wave or the spring 2011 wave through use of a computer-assisted parent interview (CAI) conducted over the telephone. The CAI was developed using globally recognized computer-assisted interviewing software that helps minimize human error during data collection (Tourangeau et al., 2017). If full demographic data was not collected in the fall of kindergarten, attempts were made at subsequent data collection time points to obtain this demographic information. The respondents were either parents or primary caregivers who identified themselves as knowing the most about the child's care, health, and education.

Sex. Within the dataset, sex is coded as either male, female, or unknown. Information concerning the child's sex was collected during the fall wave of kindergarten via parent interview using the CAI. Sex information reported during the initial wave was then compared to sex information that was gathered again in the spring of kindergarten so to confirm that the correct sex was recorded. If no response was recorded during the initial two waves of data collection, further attempts were made during the next two data collection waves to try and ensure sex information was recorded. Consequently, a child's sex information may have been collected during fall of 2011 to spring of 2013.

In efforts to aid with interpretation of the regression coefficients associated with the sex variable, the original variable in the ECLS-K data set had to be recoded. In particular, the original variable coded male participants with a 1, whereas female participants were coded as 2. At the time of the analysis, sex was recoded in such a way to make the female participants serve as the intercept, or comparison group relative to males. Thus, any regression coefficients associated with sex indicate the degree to which males deviated from females.

Race. During the CAI, parent/caregivers were asked to identify which race their child belonged to among five different categories (i.e., (1) White; (2) Black or African American; (3) Asian; (4) Native Hawaiian, or other Pacific Islander; or (5) American Indian or Alaska Native; Tourangeau et al., 2017). Subsequently, the parent/caregiver was asked whether or not their child was Hispanic. It is important to note that parent/caregivers were allowed to select more than one option for the race question; however, if the child was identified as Hispanic in the second question, then they were coded by the research team as Hispanic regardless of their racial category that was previously identified. For example, a child was categorized into one of eight categories: (1) White, not Hispanic; (2) Black or African American, not Hispanic; (3) Hispanic, race specified; (4) Hispanic, no race specified; (5) Asian, not Hispanic; (6) Native Hawaiian or other Pacific Islander, not Hispanic; (7) American Indian or Alaska Native, not Hispanic; and (8) more than one race specified, not Hispanic. The resulting composite variable is mutually-exclusive, meaning a child was only coded into one of the eight categories (Tourangeau et al., 2017).

In its raw form, the race variable within the ECLS-K dataset is uninterpretable within regression-based analyses due to it being a nominal-level variable, as regression assumes every variable to be at least at the ordinal level. Thus, four dummy variables were created to aid in later data analysis and interpretation efforts. The first dummy variable included Black, non-Hispanic participants. The second dummy variable included any participants that indicated they were Hispanic. The third dummy variable included Asian, non-Hispanic participants. Finally, the fourth dummy variable was a composite of the remaining races that participants could identify with (i.e., Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, or more than one race specified). It is important to note that with each of the dummy variables, White, Non-Hispanic participants served as the comparison group, meaning that the resulting coefficients indicate the degree to which the dummy group deviated from White, Non-Hispanic participants. Consequently, Figures 2, 3, and 5 indicate that only one race variable was utilized for the purposes of simplifying the visual model; however, the final analysis was run with all four dummy variables taking the place of the single race variable.

Socioeconomic status (SES). The authors of the ECLS-K dataset created the SES composite based on parent/caregiver's education level, parent/caregiver's occupational prestige score, and amount of household income. Given that not all respondents answered all of these questions, missing data imputation was conducted by Tourangeau et al. (2017) using the Hot Deck method. Respondents and nonrespondents were matched on characteristics such as geographic region, urbanicity, household type, age, and race. Following the matching process, missing data for each component of SES was copied from matched respondents to nonrespondents to allow for a more complete data set.

Independent variables. Assessments of EF were administered to children by trained examiners. Assessments were conducted in either an unoccupied school classroom, an unoccupied meeting room, or the school library. Arrangements were made to ensure the child was not distracted by others being assessed at the same time. Each assessment took approximately one hour to complete and occurred in the fall and spring of each grade level – kindergarten through fifth grade.

Keeping in line with the *Unity/Diversity Framework* of EF (Miyake et al., 2000), two types of data that measure EF were selected from the dataset for analysis: WM and CF. It is important to note that ECLS-K did not administer a measure of Inhibition as part of the battery of EF. Consequently, the current study was unable to analyze the degree to which Inhibition influences the constructs of interest. Given the numerous factor analysis studies that clearly demonstrate an important role of Inhibition in EF, future work in this area should make every attempt to include measures of Inhibition. Unfortunately, the use of extant data limits the researcher only to what was administered as part of the waves of data collection. Despite this weakness, it is argued that valuable information was gleamed by the analysis of WM and CF on EBP in elementary-age children.

In order to measure WM, the Numbers Reversed test of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III-COG; Woodcock, McGrew, & Mather, 2001b) was administered to participants. The WJ-III-COG is appropriate for individuals aged 2 – 90 years, and consists of 20 tests that purport to measure various cognitive abilities. These 20 tests are organized into different clusters that are related by CHC construct (e.g., Fluid Reasoning, Comprehension-Knowledge, Short-Term Memory, Long-Term Storage Retrieval, Visual Processing, & Auditory Processing). For example, the Numbers Reversed test represents a single test on the Short-Term Memory Cluster, the Working Memory Clinical Cluster, and the Cognitive Efficiency Cluster.

For measuring the construct of CF, the Dimensional Change Card Sort (DCCS; Zelazo, 2006) was utilized. The DCCS is suitable for use across the lifespan and has been established as a widely-used measure of CF. It is important to note that the DCCS has also been touted as a measure of task switching and set shifting (Zelazo et al., 2013); however, these terms can simply be viewed as synonyms of CF.

Numbers Reversed. The Numbers Reversed test of the WJ-III-COG requires participants to hold a span of verbally-presented numbers in immediate awareness and reproduce them verbally in reverse order. For example, if presented with a sequence of “6….8” then the correct response would be “8….6.” During administration of this test, children were initially given five two-number sequences. The child was required to correctly answer three consecutive two-number sequences in order to advance to the administration of five three-number sequences. This pattern of administration continued until the child was unable to answer three consecutive number sequences correctly or if they completed all the number sequences on the test (i.e., 30 items).

Three scores exist within the ECLS-K dataset for the Numbers Reversed test: a *W-score*, a standard score, and a percentile rank (Tourangeau et al., 2017). Standard scores and percentile ranks are particularly suitable for making comparisons among participants at the same time points. These scores reflect a person's normative standing relative to the sample used in the test's creation. As such, standard scores will be utilized for analysis of the path model (Figure 2). In contrary, *W-scores* are a type of standardized score that are more sensitive to growth over time. Specifically, *W-scores* are derived from a special transformation of the Rasch ability scale, representing both a person's ability and the task difficulty (Mather & Jaffe, 2016). For example, the *W*-score is centered on a value of 500 which indicates the average performance for an individual aged ten years, zero months. Thus, *W-scores* are more appropriate for longitudinal analysis since they are sensitive to growth over time and as a result, are reported for descriptive information on the participants' performance over time for the current study.

Reliability. Reliability data for Numbers Reversed was acquired through use of a split-half reliability procedure (McGrew & Woodcock, 2001). The split-half reliability procedure entails taking the odd numbered items of a test and correlating them with the even numbered items of the same test. The coefficient that is produced represents the degree to which the items are consistent among each other. Per the test battery manual (McGrew & Woodcock, 2001), the reliability coefficients ranged from 0.84 to 0.93 dependent on the age of the participant. This indicates acceptable levels of reliability for the current study as Nunnally (1978) states that a measure should demonstrate a reliability coefficient of at least r = .7 for research purposes.

Validity. Validity for Numbers Reversed was established through the forms of: (1) test content, (2) developmental patterns of scores, (3) internal structure, and (4) relationships to other external variables (McGrew & Woodcock, 2001). First, content validity was ensured through analysis of test and cluster content in relation to CHC theory. In other words, tests were created to measure a single narrow ability (e.g., lexical knowledge, language development, associative memory, visualization, working memory, etc.), whereas clusters were intended to measure a primary broad factor (e.g., Comprehension-Knowledge, Long-Term Retrieval, Visual-Spatial Thinking, etc.) within the CHC framework. The format of the WJ-III-COG is organized hierarchically, with tests fitting within clusters. This cluster concept was used in efforts to minimize the chance that test users would overgeneralize single test scores to broad, multifaceted abilities (McGrew & Woodcock, 2001).

The second form of validity measured by the test developers of the WJ-III-COG entailed the demonstration of developmental growth and decline within the cognitive abilities as a function of age. Based on the data derived from the norm group for the WJ-III-COG, the Short-Term Memory Cluster demonstrates the expected growth curve as performance increases sharply through approximately 25 years of age and then performance begins a steady decline through 85 years of age. Unfortunately, growth curve information is only available at the cluster level.

The third form of validity that McGrew and Woodcock (2001) present in the WJ-III-COG technical manual concerns internal structure. More specifically, CFA was conducted in order to test whether the WJ-III-COG exhibited a factor structure consistent with CHC theory. Specifically, the Numbers Reversed test was hypothesized to load onto a factor that McGrew and

Woodcock (2001) called Short-Term Memory, which is identical to the name they called the corresponding cluster that the Numbers Reversed test falls within. The resulting data from the CFA conducted revealed that Numbers Reversed had a 0.71 standardized factor loading on this factor. Additionally, the only other WJ-III-COG test to load onto this factor was Auditory Working Memory which had a standardized factor loading of 0.74. This provides strong evidence supporting the notion that the Numbers Reversed test is a measure of Short-Term Memory as it loaded onto the same factor with another test that purported to measure the same construct.

Finally, the last form of validity reported in the WJ-III-COG technical manual concerns the relationship the tests and clusters have with other external variables. More specifically, construct validity is reported in two forms (i.e., convergent and discriminant validity) followed by predictive validity. Convergent validity is demonstrated when a test measuring a specific construct correlates strongly with another test measuring the same or similar construct. In contrast, discriminant validity is demonstrated when a test measuring a specific construct has a low correlation with another test measuring a different construct. Lastly, predictive validity is established when a test can successfully predict a certain criterion variable such as group membership status (e.g., learning disability group vs. normal control).

Convergent validity for the Numbers Reversed test was established through a positive correlation between the WJ-III-COG Working Memory Clinical Cluster and other well-known measures of WM such as that on the Stanford-Binet Intelligence Scale, Fourth Edition (SB-IV; Thorndike, Hagen, & Sattler, 1986). For example, the Working Memory Clinical Cluster of the WJ-III-COG demonstrated a correlation of $r = .64$ with the Short-Term Memory Composite on the SB-IV. Unfortunately, the technical manual offers no specific data on the relationship between Numbers Reversed and other tests in the field; only cluster information is available.

Discriminant evidence of Numbers Reversed on the WJ-III-COG was established through weak correlations between this test and other well-established measures that do not intend to measure WM. For example, a correlation of $r = .18$ was reported between Numbers Reversed and both the Verbal Comprehension Index and Perceptual Organization Index of the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III; Wechsler, 1997). This relationship is far from a perfect correlation of $r = 1.0$ indicating that the Numbers Reversed test is measuring something clearly distinct from that measured by the Verbal Comprehension Index and the Perceptual Organization Index on the WAIS-III.

Finally, predictive validity was established through analyzing the correlations between the Working Memory Clinical Cluster on the WJ-III-COG and other constructs it should be able to predict such as academic achievement and group membership status. For example, the manual indicates that the Working Memory Clinical Cluster had strong correlations with the Total Achievement Cluster of the WJ-III Tests of Achievement (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001a). This means that high scores on the Working Memory Clinical Cluster predicted high scores on the Total Achievement cluster on the WJ-III-ACH. Further, the manual (McGrew & Woodcock, 2001) also indicates that the Working Memory Clinical Cluster was correlated to reading ($r = .37$) and spelling skills ($r = .34$) as measured by the Wide Range Achievement Test, Third Edition (WRAT-III; Wilkinson, 1993).

Regarding group membership status, the manual (McGrew & Woodcock, 2001) demonstrates that the Working Memory Clinical Cluster successfully distinguished between students with learning disabilities and students without learning disabilities as evidenced by significant mean differences on the Working Memory Clinical Cluster between groups. More specifically, students without a learning disability (M = 105.7; SD = 12.1) scored at least 10 standard score points higher, on average, relative to students with a learning disability (M = 94.7; SD = 13.5). This finding is powerful evidence supporting the validity of WJ-III-COG to measure WM given that the cluster was able to distinguish between students with a learning disability and those without.

Dimensional Change Card Sort. On the standard version of the DCCS (Zelazo, 2006), children are presented with two target cards (i.e., a blue rabbit and a red boat) and are then subsequently asked to sort 14 cards of either red rabbits (n = 7) or blue boats (n = 7) into piles based on a single dimension (i.e., color). After completion of this first phase (i.e., pre-switch phase), the children are asked to sort the same test cards based upon another dimension (i.e., shape), with this phase being called the post-switch phase. During the preschool years (i.e., age three through five years), success on this measure is calculated by the number of items correctly sorted during the post-switch phase; with the majority of three-year-olds unable to sort by a second dimension (Cragg & Chevalier, 2012; Zelazo, 2006).

During data collection for the kindergarten through first grade, children were administered the standard version of the DCCS with some minor variations in efforts to expedite test administration for such a large sample of children. The children were asked to sort a series of 22 picture cards (11 for each dimension) according to various rule sets using these same cards throughout each sorting task. First, during the Color Game, children were asked to sort cards into two trays (i.e., red rabbit or a blue boat) according to color. Following the color game, children were asked to play the Shape Game which required the child to place rabbits into one pile and boats into another pile. If the child was able to sort four out of six cards during the

Shape Game, he/she would move on to a third game called the Border Game. The Border Game required the child to sort by either color or shape depending on whether or not the card had a black border around the edges (i.e., if the card had a border then the child sorted by color; if no border was present then the child sorted by shape). The difficulty in the Border Game lies in the child's ability to switch between sorting by color or shape as determined by whether the card has a black border around the edges.

Level of performance on the DCCS (Zelazo, 2006) for grades K-1st were measured using a standard assessment protocol with three raw scores being yielded (range = 0-6) and one total score derived from the summing of the raw scores (range = 0-18). First, a pre-switch score was calculated indicating the number of cards the child correctly sorted by color. For example, a score of 6 would indicate that the child correctly sorted 6 out 6 cards that were administered. Following the pre-switch trials, children were administered the post-switch trials where they were instructed to sort by shape. If the child scored at least four items correct on the Shape Game, the Border Game was administered and scored. Performance on the Border Game had a possible range of 0 - 6 and required the child to determine the sorting rule based upon whether or not the card had a black border present along the edges. The number of cards the child correctly sorted when the sorting rule was applied by the presence of a black border on the edge of the cards made up the Border Game score (Tourangeau et al., 2017). Given that not all children scored well enough on the Shape Game (i.e., at least a score of 4/6) to proceed to the Border Game, a score of "Not Administered" was coded for some participants.

Information from the ECLS-K manual indicates that post-switch scores for the sample were relatively high (e.g., weighted mean during fall of kindergarten = 5.23; SD = 1.679), meaning that a large portion of children did well on this measure. Based upon this preliminary finding, the authors (Tourangeau et al., 2017) recommend that researchers utilizing the K-1st grade data should use the composite score that involves summing the pre-switch score, the post-switch score, and the Border Game score. As such, the current study followed the authors' recommendations by using this composite variable and converting it to a *z-score* in order to have a standardized score for all analyses. For children who were not administered the Border Game Score, a score of 0 was added to their pre-switch and post-switch score.

For second grade, both administration and scoring changed for the DCCS in order to make the assessment more age-appropriate given that such high numbers of participants were reaching ceiling-level performance (Tourangeau et al., 2017). In particular, beginning in the fall of second grade, administration of the DCCS switched to a computerized format where the cards were presented on a computer screen and children sorted them into virtual piles. Following the presentation of a female voice that instructed what dimension to sort by, the children were instructed to use specific keys on the keyboard in order to indicate where they wanted to sort the cards. In conjunction with switching to a computerized format, children in second grade also began the measure at a later starting point. Following the completion of no less than eight and no more than 24 practice trials, children who demonstrated understanding of the task began the test. Instead of being administered five pre-switch trials and five post-switch trials where children were asked to sort by one dimension, children in second grade began immediately on the 30 mixed-block trials in which the sorting rule varied by trial. Consequently, children in second grade were given credit for completing the pre-switch and post-switch trials with 100% accuracy.

During the mixed-block trials, the sorting rules were mixed randomly; however, each participant was required to sort by one rule more than the other. More specifically, every child was asked to sort by shape for 23 trials and color for seven trials. Again, although this ratio was fixed for every child, the order in which the sorting rules were presented was randomized. The administration of more shape trials was done in order to build a response tendency (Tourangeau et al., 2017). According to the author of the DCCS (Zelazo et al., 2013), the building of a response tendency makes it harder for children to inhibit the dominant response in favor of the non-dominant response. As such, accuracy and reaction times for the trials where the nondominant response is required typically demonstrate the cost associated with shifting to the nondominant response. In other words, older participants typically slow down on nondominant trials so to preserve accuracy at the expense of speed; however, younger participants (e.g., under the age of eight years) typically respond impulsively and less accurately (Zelazo et al., 2013).

In addition to the administration change that occurred in second grade, scoring protocol for the DCCS also changed in order to address the lack of variability that was present in the data. In particular, computer administration of the DCCS allowed for the inclusion of reaction time in the calculation of the final score. The inclusion of reaction time provides valuable information about the student's performance because older children (i.e., over the age of eight years) are able to accurately sort with ease (Tourangeau et al., 2017; Zelazo et al., 2013). Since most children eight years and older can sort accurately, the inclusion of reaction time provides another measure of performance, allowing the researcher to detect differences between participants based on how fast they sort the cards.

As already stated, the overall score for the second-grade administration of the DCCS is calculated based on the students' accuracy and reaction time. In particular, total scores range from zero to ten, with differential weight given to accuracy (0-5) and reaction time (0-5). According to the manual (Tourangeau et al., 2017), accuracy is accounted for first as reaction time is only included in the child's total score if their accuracy was 80% or higher. Consequently, if the child was 80% accurate or better then reaction time was also included in the final calculation of their total score. Of the administered trials, a child was given .125 points per correct response, leaving a possible 40 accuracy points that could be accrued. For the calculation of the reaction time points, the child's median reaction time to correct nondominant trials was used. The minimum median reaction time that was allowed was 500 milliseconds and the maximum was 3000 milliseconds. Median reaction times outside of this range were set at the threshold that was exceeded. Finally, in order to create a more normal distribution of reaction time scores, the ELCS-K research team (Tourangeau et al., 2017) performed a log (base 10) transformation to the median reaction time scores. The resulting score ranges from zero to five, with higher scores representing faster reaction times. The formula that was used is presented below:

$$Reaction\ time\ score = 5 - (5 * \left[\frac{\log RT - \log(500)}{\log(3000) - \log(500)}\right]$$

Unfortunately, given the change in scoring from 1st to 2nd grade, comparison between participants CF scores on the physical and electronic versions was not possible since the total score that was calculated is different. As such, the current study was only able to detect growth in CF from K-1st grade. Despite this loss, the ECLS-K panel members indicated that the results should be comparable across the different administration formats and that converting raw scores to standardized scores should facilitate any comparisons that researchers utilizing this data want to make (Tourangeau et al., 2017). Consequently, the current study converted the raw score performances from K-1st and 2nd grade into *z-scores* based off the weighted means and standard deviations that are provided within the dataset.

Reliability. Zelazo and colleagues (2013) demonstrated excellent test-retest reliability (Intraclass correlation [ICC] = .92) during their validation study of the measure which included a two-week interval between testing. This indicates that the DCCS exhibits temporal stability meaning the researcher can be reasonably assured that the trait they are assessing (i.e., CF) is fairly stable over time.

Validity. Validity for the measure was established through use of convergent and discriminant correlational methods. Convergent validity was established by examining the relationships between the DCCS and existing measures of EF (Zelazo et al., 2013). In particular, the Block Design subtest of the Wechsler Preschool and Primary Scale of Intelligence, 3rd Edition (WPPSI-III; Wechsler, 2002) was used for a three to six-year-old group; whereas the Color-Word Interference Inhibition subtest of the D-KEFS (Delis, Kaplan, & Kramer, 2001) was used for the 8-year-old to 15-year-old group. The rationale behind including Block Design as a convergent measure for EF was established through prior research indicating a strong correlation between fluid reasoning and EF (Blair, 2006).

The results of the validation study indicated a strong correlation (r = .69; p < .0001) between the DCCS and the Block Design subtest of the WPPSI-III, providing strong evidence of convergent validity for the three to six-year-old group. Similarly, the correlation between the DCCS and the Color-Word Interference Inhibition subtest of the D-KEFS was also strong (r = .64; p < .001), providing further evidence of convergent validity for the eight through 15-year-old group. Taken together, both results provided enough evidence to justify use of the DCCS with the current sample.

Discriminant validity was established utilizing the Peabody Picture Vocabulary Test, 4th Edition (PPTV-IV; Dunn & Dunn, 2007). The PPTV-IV is a well-established measure of receptive vocabulary, suitable for use with individuals aged three-85 years. Interestingly, high correlations between the PPTV-IV and DCCS were found for the younger children (i.e., three through six years of age) ($r = .79$; $p < .001$); however, these correlations declined with the older group (eight through 15 years of age) ($r = .55$; $p < .001$). One potential reason for why the relationship between the DCCS and the PPTV-IV was stronger for the younger group relative to the older group is the degree to which EF was recruited for the differing age groups. More specifically, the younger children likely had to rely on EF more as their verbal skills are not quite as developed. The authors support this viewpoint, stating that this decline in the strength of the correlation as a function of age is likely due to increased differentiation between EF and receptive vocabulary ability later in life (Zelazo et al., 2013). Despite this finding, a lower correlation would be preferred to establish discriminant validity; therefore, more research is needed in the area of discriminant validity for the DCCS.

Dependent variable. Data on the magnitude of EBP was collected through use of the Social Skills Rating System - Teacher Rating Form (SSRS-TRF; Gresham & Elliott, 1990). On the SSRS-TRF, composite scores are formulated from scores on various subscales. In particular, the Externalizing Problem scale of the Problem Behaviors Composite was utilized. As per the manual for the SSRS, EBP are defined as “inappropriate behaviors involving verbal or physical aggression toward others, poor control of temper, and arguing” (Gresham & Elliott, 1990, p. 4). The SSRS measures the frequency in which these EBP occur through a 3-point Likert-type scale (i.e., Never, Sometimes, Very Often).

In efforts to expedite indirect assessment of the child participants, the ECLS-K research team utilized an abbreviated version of the SSRS-TRF. In particular, six items were administered with these six items either being directly taken from the original SSRS-TRF

(Gresham & Elliott, 1990) or slightly adjusted for the ECLS-K study (Tourangeau et al., 2017). Data about the specific items that were used was unavailable due to copyright restrictions. In order to obtain a score for EBP, at least 4 out of the 6 items that were used must have been answered by the teacher who was rating the child. The manual indicates that similar to the original SSRS-TRF (Gresham & Elliott, 1990), higher scores indicate the presence of more frequent EBP. The scores within the dataset range from 1-4. For the purposes of making comparisons to other variables, the current study converted the SSRS-TRF score (e.g., 0-4) to z-scores based off the reported mean and standard deviation for each grade level.

In general, research has found a low level of agreement between teacher and parent ratings of EBP (Korsch & Petermann, 2013), meaning that teachers and parents tend to disagree on the degree to which a child exhibits acting-out, problem behaviors. It is important to note that since teachers were utilized as informants, scores on this scale indicate the degree to which the child exhibits these acting-out behaviors in the classroom. In contrast, parents were asked to rate how the child behaves in the home and community, as it is assumed that parents are the expert raters in these contexts. When comparing the ratings of EBP between teachers and parents, prior research has found that teachers tend to be more accurate raters over time (Verhulst, Koot, & Van Der Ende, 1994). This latter finding, in conjunction with a desire to focus on children's problem behaviors in the academic context, is why this study employed teacher ratings as a measure of EBP.

Reliability. Three methods of establishing reliability were utilized in the validation of the SSRS-TRF: internal consistency, test-retest, and interrater. For elementary aged children, the coefficient alpha for the EBP scale was .88, indicating an acceptable level of internal consistency (Nunnally, 1978). Test-retest reliability also indicated temporal stability (four weeks between both administrations) for the TRF as the coefficient yielded during norming was *r* = .82. Finally, interrater reliability between teacher and parent ratings of the EBP subscale was found with a *r* = .27 for elementary-aged children. Although this correlation is low, it is important to note that parents and teachers were asked to rate the child in different contexts. For example, parents may experience different behaviors from their child at home relative to what the child's teacher experiences in school given the differing demands for each environment. In summary, the SSRS has adequate levels of reliability to address the current research questions.

Validity. Validity for the SSRS was also established using multiple methods: content validity, criterion-related validity, convergent validity, and factor analysis. First, the manual states (Gresham & Elliott, 1990) that the SSRS was developed based on a broad survey of the empirical literature, with this indicating that the test items represent the domains they purport to measure – this serves as evidence of the content validity for the measure (Gresham & Elliott, 1990). Second, criterion-related validity was established using measures that depicted outcomes the authors believed the SSRS should be able to predict. For the TRF, three separate validity studies indicated moderate to high correlations (range: *r* = 0.41 – *r* = .69) between the EBP subscale of the SSRS and other well-established measures for problem behaviors (e.g., Social Behavior Assessment; Stephens, 1981; Child Behavior Checklist-Teacher Report; Achenbach & Edelbrock, 1983; Harter Teacher Rating Scale; Harter, 1985, respectively); with this providing evidence of the SSRS's validity for the elementary-age population.

Third, convergent validity was found between other well-established measures of EBP and the TRF of the SSRS, with correlation coefficients ranging from moderate to high. For example, a correlation of *r* = .81 was found between the TRF and the CBCL Teacher Total Behavior Problems score (Achenbach & Edelbrock, 1983). In addition, a correlation of *r* = .55

was found between the TRF and the Total score on the Social Behavior Assessment (SBA) (Stephens, 1981); with high scores on the SBA also indicating behavior problems. Taken together, both of these examples offer support for the convergent validity of the TRF on the SSRS.

Finally, factor analysis of the elementary data for the EBP subscale yielded factor loadings for the specific items ranging from .59 to .86, indicating strong correlations among the items that purport to measure EBP. In other words, all of the items on the EBP subscale loaded onto a single factor, justifying the TRF's EBP subscale as a distinct construct from the other items on the TRF. Unfortunately, the manual does not provide the eigenvalues for each factor so it is impossible to determine whether the EBP subscale/factor had an eigenvalue greater than one (i.e., Kaiser's rule; Kaiser, 1960); however, it is worth mentioning that the authors state that they only retained factors that met Kaiser's rule. In summary, the TRF of the SSRS has multiple sources of evidence that establish acceptable validity for the current study's purposes.

Data Preparation

This section delineates how the current study addressed two issues associated with using complex survey data. The first issue that was addressed involved the sampling procedure utilized in order to ensure adequate representation of certain racial groups within the final sample. The second issue that was addressed involved how missing data was managed in the data analysis.

Weighting and jackknife replication variance estimation. The sampling procedure utilized by the ECLS-K dataset incorporated a clustered design in which specific geographic regions were oversampled in efforts to achieve adequate minority representation. As such, in its' raw form, the ECLS-K sample is disproportionately represented by those from Asian, Native

Hawaiian, and other Pacific Islander (API) descent. Similarly, given the clustered sample design that grouped people by counties and then schools, participants may be more similar than intended (i.e., people from a given area may inadvertently share characteristics as a function of their geographic location), leading to biases inherent in subsequent estimations of variance and standard error (Thomas & Heck, 2001). The problems of oversampling and homogeneity within sampled clusters had to be addressed in order to ensure the sample used in the current study was truly representative of the national distribution.

To solve these problems, the current study employed sample weights that were derived from the ECLS-K manual (Tourangeau et al., 2017) in order to restore balance to the sample given that those from API descent were overrepresented. Providing corrective weighting to the sample ensured that subsequent analyses were conducted on a truly representative sample as opposed to one disproportionately represented by the oversampled groups. Weights are applied to datasets using commands embedded in most statistical programs such as the Statistical Package for the Social Sciences (SPSS).

In addition to applying sample weights, jackknife replication variance estimation was used to adjust standard errors so subsequent analyses that relied on standard errors were accurate. In other words, the jackknife procedure addressed the homogeneity that is present as a result of the clustered-sample design employed by the ECLS-K dataset. This helped diffuse the influence that a lack of independence among observations may have had on the dataset. Again, this was conducted using commands within the statistical program of choice.

Missing data. Missing data often plagues large public-use data sets, leading the researcher to drop cases that have incomplete information. This can be particularly problematic to longitudinal data that follows the same participants over an extended time period (Young &

Johnson, 2015). Further convoluting this problem of missing data, systematic patterns may exist as to why a participant has missing data leading to a potentially biased sample (e.g., nonrandom item nonresponse & nonrandom sample attrition). In the current study, Rubin's (1996) method of multiple imputation was utilized by the ECLs-K research team for the race and SES variables (Tourangeau et al., 2017). No other variables required multiple imputation as the sample size remained large enough for the current purposes. Multiple Imputation involves imputing plausible values in place of the missing values based on Bayesian modeling (Rubin, 1996).

Analyses

To address the research hypotheses, the current study employed a within-subjects correlational design, utilizing path analysis to analyze longitudinal data from kindergarten through second grade. Path analysis is a form of structural equation modeling (SEM) that utilizes observed variables, as opposed to latent variables which are not directly measured, in efforts to test a causal model (Keith, 2014). This statistical technique has numerous advantages such as allowing the researcher to measure a hypothesized causal influence, as well as compare the magnitude of influence that independent variables have on a given dependent variable. Additionally, path analysis also allows researchers to test complex pathways in which one independent variable works through another independent variable to impact the dependent variable. Thus, path analysis facilitated additional understanding of the causal relationship between EF and EBP, as well as the developmental relationship between WM and CF.

The hypothesized model is presented in Figure 2. The model is recursive meaning that the direction of influences flows in a single direction and not vice versa (Acock, 2013). Curved lines are drawn to represent covariance or correlations between variables. Single-headed lines depict the hypothesized direction of causal influence between two variables (e.g., WM to CF, CF to EBP). Disturbances, or unexplained variances, are also depicted in the path model and represent the amount of variance that is left unexplained by the predictor variable(s) used for a given endogenous variable, or variable that is predicted by another. Given that observed variables, rather than latent variables, were utilized in this model, all variables are represented by rectangles which indicate that they are manifest, or observed variables.

In path analysis, three types of variables are typically identified: exogenous, endogenous outcome, and endogenous mediator variables (Acock, 2013). Exogenous variables are variables that have no explanatory variables before them. In the current study, the control variables of Sex, Race/Ethnicity, and SES were the exogenous variables. The rationale behind including these variables within the model was based on prior research (Buckhalt, El-Sheikh, & Keller, 2007; Deater-Deckard, Dodge, Bates, & Pettit, 1996; Deater-Deckard, Dodge, Bates, & Pettit, 1998; Dodge, Pettit, & Bates, 1994; Mezzacappa, 2004; Nesbitt, Baker-Ward, & Willoughby, 2013) that has demonstrated a link between these variables (i.e., sex, race, and SES) and variables of interest in the current study (i.e., WM, CF, and EBP).

The endogenous outcome variable of the current study included the spring of second grade measure of EBP. This variable is represented by a rectangle with one-way arrows pointing toward it from other independent variables (see Figure 2). The graphic illustration indicates that the dependent variable is influenced by other preceding variables in a hypothesized causal fashion. Again, causation cannot be proven with correlational methods; however, evidence suggesting causation can be acquired using path analysis.

Finally, the endogenous mediator variables are represented by variables that have both one-way arrows pointed towards them and one-way arrows pointed away from them. In other words, endogenous mediator variables are influenced by preceding variables while also simultaneously influencing subsequent variables. An example in the current model are all the time points of CF as these measures have arrows from earlier time points of WM or one of the control variables drawn towards them, while simultaneously having arrows pointed towards the dependent variable (i.e., EBP). This means that the current model hypothesized that CF mediates that relationship between WM and EBP. Another example of an endogenous mediator variable within the current study's model is all time points of WM as well, as the control variables have arrows pointed towards all measures of WM, with WM having arrows pointed towards EBP.

Model. The current model was derived from logic, review of the literature, and theory. Further, given the longitudinal nature of the extant data, temporal ordering of the variables was utilized to further strengthen the argument of a causal relationship between the variables of interest. Despite relying on prior research, theory, and temporal ordering, the current study was not experimental, and therefore, the resulting model did allow for definitive causal statements to be made. As with all correlational methods, there exists the third-variable problem, which states that there can always be another confounding variable explaining the relationship that is observed between any two given variables. For example, one may observe a significant correlation between EF and EBP; however, this does not mean that the relationship is not explained by another variable that was not measured.

Working from left to right in the hypothesized model illustrated in Figure 2, the curved line between race and SES is justified by prior research (Shifrer, Muller, & Callahan, 2016; Williams, 1999) that has demonstrated overrepresentation of minorities in lower SES groups. Next, straight lines from sex, race, and SES to WM1 and CF1 are justified by prior research that has suggested a relationship between sex, familial background characteristics and EF (Buckhalt, et al., 2007; Mezzacappa, 2004; Nesbitt et al., 2013; Raver et al., 2013; Sarsour et al., 2011;

Ursache et al., 2015; Voyer et al., 2016). Subsequent arrows to later time points of WM and CF were not needed since this data was longitudinal in design and was measuring the same constructs over time. Straight lines from race, and SES were also drawn to EBP. Justification for these hypothesized relationships also has support from prior research investigating these variables (Deater-Deckard et al., 1998; Dodge et al., 1994), suggesting that race and SES significantly influence the presence of EBP through various pathways such as parenting practices and exposure to the accumulation of risk factors associated with living in poverty.

The next part of the model included a temporal ordering of WM and CF from the fall of kindergarten year to the spring of second grade from left to right. This panel design allowed the current study to make conclusions of causality as the direction of influence was clearly delineated by time points. With that, arrows from earlier time points of each WM and CF component are drawn to later time points of each respective component, as it is thought that early demonstration of WM and CF sets the foundation for subsequent development of each construct, respectively. Next, straight arrows from earlier WM time points are drawn to subsequent time points of CF (e.g., an arrow from WM1 in the fall of Kindergarten to all subsequent time points of CF [CF2, CF3, CF4, CF5, CF6]). As already stated, one of the main hypotheses of the current study was that WM influences the development of CF, while simultaneously having direct and indirect effects on EBP. For example, an individual with high WM in kindergarten would have earlier development of their CF as the current project hypothesized that CF is built off of successful WM. Further, an individual high in both WM and CF was hypothesized to demonstrate lower levels of EBP.

To represent a correlational relationship, curved lines are drawn between concurrent time points of each EF component (e.g., WM1 and CF1, WM2 and CF2, etc.). This was done in order to emphasize that the two components of EF were related, but with no known direction of influence since the two are measured at the same time point. Finally, each time point of WM and CF (e.g., WM1, CF1, WM2, CF2, etc.) have arrows drawn to EBP. The rationale behind inclusion of these arrows was based on prior research demonstrating a relationship between EF deficits and various externalizing issues such as ADHD and EBP as early as the preschool years (e.g., Antshel et al., 2014; Barkley, 2014; Biederman et al., 2004, Schoemaker et al., 2013). Inclusion of all of the time points also allowed the current study to determine the extent to which deficits in either component of EF predicted EBP in second grade, if at all. It was hypothesized that WM and CF, as measured at kindergarten, would be a significant predictor of EBP; with individuals who are low on EF exhibiting higher degrees of EBP. Further, the relationship between WM/CF and EBP was hypothesized to increase as the measure of WM/CF became temporally closer to the later measure of EBP.

Chapter IV: Results

Given the longitudinal nature of the current study, numerous cases from the base year sample of 18,174 were missing relevant data for at least one of the variables of interest by second grade (i.e., race, sex, WM, & CF), with exception of the SES variable which underwent multiple imputation to account for nonresponse. As a result, 5,711 participants were included in the final analysis, with this resulting sample not being representative of second grade children in the year of 2012-2013. Despite not having a nationally representative sample of second graders, the use of the second-grade weight variable provided corrected estimates that are representative of kindergarteners in the 2010-2011 cohort. Thus, the following findings are based off of a nationally representative sample of kindergarteners from the 2010-2011 cohort who progressed through second grade, with complete data for all the variables included in the current study's model. Please refer to Table 1 for information pertaining to the sample of base-year respondents, but remember, this chart is not representative of the final sample that was utilized. Unfortunately, descriptive information relating to race and sex was unable to be collected due to the use of sampling weights and jackknife variance estimation.

Table 2 provides descriptive information for all the variables included in the model. This is where you will find means, standard errors, and confidence intervals for each of the variables that were included in the study. *W*-Scores are included for the measure of WM (i.e., Numbers Reversed) for the purposes of visually inspecting growth from time point to time point; however, please note that standard scores were used in the analysis. Finally, Tables 3, 4, and 5 provide a summary of the direct, indirect, and total effects yielded from the current model, with appropriate adjustments made via sample weighting and jackknife variance estimation. Given the choice of using path analysis for testing the hypotheses of the current study, regression coefficients are reported in unstandardized and standardized formats (*β*). Standardized regression coefficients represent change in standard deviation units (*z*-scores) within the dependent variable of interest and can be compared between variables to determine the relative contribution each variable makes to a given dependent variable. As for determining the magnitude of effects, Keith's (2014) guidelines were used, which specify: *β's* above .05 are considered small, *β*'s above .10 are considered moderate, and *β*'s above .25 are considered large. In addition to coefficients, *t*-scores and their associated *p*-values are reported in order to highlight which paths yielded significant results. Unfortunately, as a byproduct of using complex survey data, no goodness-of-fit statistics were able to be calculated.

As regression coefficients are derived from an estimated covariance matrix within path analysis, standardized covariances are also presented via Pearson's *r* for the curved lines specified by the model (i.e., the full correlation matrix is unable to be calculated given the use of sample weighting and jackknife variance estimation; see Figure 4 and Figure 5). These values range from $r = -1.00$ to 1.00, with a negative sign indicating an inverse relationship while a positive sign indicates that both variables move in the same direction. As for determining the strength of correlations, the classification scheme offered by Dancey and Reidy (2004, as cited in Akoglu, 2018) was utilized. Correlations ranging from $r = \pm 0.00$ to 0.3 are considered weak; correlations ranging from $r = \pm 0.4$ to 0.6 are considered moderate; and correlations ranging from $r = \pm 0.7$ to 0.9 are considered strong.

Analysis Summary

Overall, the path model yielded mixed results regarding support for the aforementioned hypotheses. In particular, aside from the control variables (sex, race, & SES), fall of kindergarten CF (CF1) and fall of second grade WM (WM5) were the only variables to have a significant direct effect on the spring of second grade EBP. In addition, results failed to support the hypothesis that CF would significantly mediate the relationship between WM and EBP, as no significant direct effect of CF2 on EBP was found. Finally, results failed to support the hypothesis that the relationship between both WM and CF to EBP would be strongest for the temporally closer measures of WM and CF. Analysis of standardized coefficients for WM did demonstrate this pattern, with WM5 showing the highest β value; however, the opposite pattern was true of CF. In particular, the strongest time point for predicting later EBP was fall of kindergarten (i.e., CF1), directly contradicting the original hypothesis.

Despite failing to reject the null hypothesis for all of the hypothesized relationships, the current study did find supporting evidence for the theory that WM influences the development of CF. In fact, WM1 significantly impacted CF scores through the fall of second grade, highlighting that a child's WM in the fall of kindergarten can be used to predict fall of second grade CF scores to a small, yet meaningful degree. Similarly, fall measures of WM were found to significantly predict spring measures of CF for each grade level. In sum, WM not only showed predictive power for CF from fall to spring in each respective grade level, but WM also predicted CF across multiple grade levels. Consequently, this is one of the first studies to analyze the contribution of WM to CF across multiple grade levels, while controlling for key demographic variables (i.e., sex, race, and SES), with a nationally-representative sample of kindergarten students.

Findings Related to Control Variables

Direct, indirect, and total effects were estimated for the following control variables in order to parse out the influence that these common causes have on the independent variables (WM and CF) of the current study, as well as on the dependent variable (EBP). Direct effects represent the direct influence a given independent variable has on a dependent variable, whereas indirect effects represent the influence the independent variable has on a dependent variable through the independent variable's influence on mediating variables. Finally, the total effect represents the sum of the direct and indirect influence a variable has on the dependent variable of choice. Please see Figure 4 for the correlational findings among the control variables. As for the findings related to direct and indirect effects, please see Tables 3, 4, and 5.

Sex. The direct effect of sex on WM, as measured in the fall of kindergarten (WM1), yielded a small effect significant at the $p = .042$ level ($\beta = -.04$; $t = -2.10$) indicating that males and females performed significantly different on the Numbers Reversed test of the WJ-III-COG. In particular, males scored .04 standard deviations less, on average, than females during the fall of kindergarten. Similarly, sex had a small, yet significant direct effect ($\beta = -.07$; $t = -2.39$; $p = .022$) on CF1, indicating that males and females differed in their level of CF at the onset of kindergarten as well; with females outscoring males yet again by an average of .07 standard deviation units. Furthermore, the direct effect of sex on EBP yielded a moderate, significant relationship ($\beta = .14$; $t = 5.53$; $p < .001$), indicating that males and females differed significantly in the number of symptoms of EBP their teachers endorsed; with males showing .14 standard deviations more EBP than females, on average, during the spring of second grade. Finally, no significant indirect effect was found between sex and EBP ($\beta = .00$; $t = 1.27$; $p = 0.213$), although the total effect of sex on EBP was moderate and significant at the $p < .001$ level ($\beta = .14$; $t = 5.97$). In sum, females generally showed higher levels of EF at school entry, and lower levels EBP during the spring of second grade.

Race. As already mentioned, four dummy variables were utilized during the path analysis in order to yield interpretable coefficients that reflected the degree to which the specified racial group(s) differed from White, non-Hispanic participants. First, Black, non-Hispanic participants (β = -.13; *t* = -3.96; *p* < .001), as well as Asian, non-Hispanic participants (β = -.07; *t* = -3.53; *p* < .001), demonstrated significantly lower levels of WM on average relative to White, non-Hispanic participants during the fall of kindergarten. In contrast, Hispanic participants (β = -.10; *t* = -1.94; *p* = .060) and participants from the other race category (i.e., Native Hawaiian or other Pacific Islander, American Indian or Alaska Native, or more than one race specified; β = -.03; *t* = -1.29; *p* = .205) did not score significantly different from the White comparison group on WM1.

In regard to presenting levels of CF during the fall of kindergarten, a similar pattern emerged. First, Black (β = -.08; *t* = -2.50; *p* = .017), Hispanic (β = -.13; *t* = -3.78; *p* = .001), and Asian (β = -.10; *t* = -4.04; *p* < .001) participants all scored significantly lower than the White comparison group; with exception to the other race group which scored lower than White participants, but the difference did not reach significant levels (β = -.02; *t* = -0.69; *p* = .493). As for race's impact on EBP, significant differences between White participants and the other racial groups were found for two groups. More specifically, both Black (β = .04; *t* = 2.24; *p* = .031) and Hispanic (β = -.06; *t* = -2.64; *p* = .012) participants showed significantly different levels of EBP relative to Whites. A closer look at the coefficients reveals a different pattern, however, as Black participants demonstrated .04 SD units more EBP than White participants on average; whereas Hispanic participants exhibited -.06 SD units fewer EBP on average. As for Asian participants (β = -.06; *t* = -1.28; *p* = .209) and participants from the other race category (β = .01; *t* = 0.57; *p* = .574), no significant differences in EBP were detected during the spring of second grade.

Analysis of the indirect effects of each racial category on EBP yielded interesting findings as well. These relationships specify the degree to which a person's racial category influenced their later EBP through WM and CF relative to White participants. In total, no significant indirect effects were found among the racial groups. This provides evidence in support of the notion that regardless of a person's racial group, a similar pattern between EF and EBP is found. As for the total effects of a person's racial group on EBP, or the summation of direct and indirect effects, significant total effects were found with Black ($\beta = .04$; $t = 2.04$; $p = .048$) and Hispanic participants ($\beta = -.06$; $t = -2.47$; $p = 0.018$); however, no significant differences were found between White participants and participants identified as Asian ($\beta = -.06$; $t = -1.07$; $p = 290$) or from the other race category ($\beta = .01$; $t = 0.64$; $p = .524$).

In terms of correlational data, numerous significant relationships were noted. First, a weak, but significant relationship between Black participants and SES was found ($r = -.14$; $p = .016$). In addition, Hispanic participants also revealed a weak, yet significant negative relationship with SES ($r = -.35$; $p < .001$). No other significant relationships were found between race and SES. As for the relationship between race and sex, only those from Asian descent demonstrated a significant relationship with sex ($r = -.046$; $p = .024$), suggesting a small bias within the Asian sample towards a male identification. No other notable correlations were observed between race and the other control variables.

SES. Paths from SES to WM1 ($\beta = .18$; $t = 10.44$; $p < .001$), CF1 ($\beta = .09$; $t = 3.95$; $p < .001$), and EBP ($\beta = -.06$; $t = -3.20$; $p = .003$) all yielded statistically significant results. As for the direction of the relationships, the general pattern observed is that as an individual's SES increased, their EF increased and the presence of teacher-endorsed EBP decreased. In fact, SES predicted a child's WM1 with a moderate strength relationship, a child's CF1 with a small to moderate relationship, and the presence of EBP with a small relationship. Analysis of the indirect effects of SES on EBP failed to reach significance (β = -.00; *t* = -0.14; *p* = .887); however, the total effect of SES on EBP was found to be a small, yet significant relationship (β = -.06; *t* = -2.25; *p* = .030). Thus, SES shows predictive strength for determining a child's EF and the degree to which they will experience problematic behaviors in the spring of second grade.

Main Findings Related to Hypotheses

Hypothesis 1. In the present study, it was hypothesized that WM1 would have a significant negative relationship with the presence of EBP in the spring of second grade (See Figure 6). Data yielded from the path analysis failed to support this hypothesis as the path leading from WM1 to EBP failed to reach significance (β = .06; *t* = 0.84; *p* = 0.406). Despite the direct effect failing to reach significant levels, the indirect effects of WM1 on EBP through later WM and CF did reach significance (β = -.04; *t* = -2.82; *p* = 0.007), indicating that WM1 influences EBP to a small, but meaningful degree, primarily through its influence on other variables such as later time points of WM and CF. Convergent with the lack of a significant direct effect of WM1 on EBP, paths from WM, as measured in the spring of kindergarten (WM2; β = -.03; *t* = -0.99; *p* = 0.326), and the fall (WM3; β = -.04; *t* = -0.73; *p* = 0.472) and spring of first grade (WM4; β = .02; *t* = 0.53; *p* = 0.600), all failed to reach significance. Again, despite no significant direct effects being noted for these time points, WM4 did significantly impact EBP indirectly to a small degree (β = -.03; *t* = -2.85; *p* = 0.007), with WM2 (β = -.04; *t* = -1.97; *p* = .056) and WM3 (β = -.02; *t* = -1.82; *p* = .076) yielding measurable indirect effects but failing to reach significant levels. In contrast to the other time points of WM, fall of second grade WM (WM5) was found to have a small, yet significant direct effect on spring of second grade EBP (β = -.07; *t* = -3.52; *p* = 0.001). More specifically, for every standard deviation unit increase in

WM5, participants exhibited -.07 standard deviation units less of EBP. Finally, the correlation between spring of second grade WM (WM6) and EBP was nonsignificant (r = -.018; p = 0.471).

Hypothesis 2. In the present study, it was hypothesized that CF1 would have a significant negative relationship with the presence of EBP in the spring of second grade (See Figure 6). Results from the path model rejected the null hypothesis as a small, yet significant direct effect between CF1 and EBP was found (β = -.06; t = -2.03; p = 0.049). More specifically, for every standard deviation unit increase in CF, participants exhibited -.06 standard deviations fewer EBP. In contrast, the indirect effect of CF1 on EBP failed to reach significance (β = -.00; t = -0.43; p = 0.668); however, significance testing of the total effect did yield significant results (β = -.07; t = -2.31; p = 0.026). Thus, when taking into account both the direct and indirect effects that CF1 have on EBP, results indicate a small, but meaningful predictive influence on later EBP, primarily through the direct effect. Analysis of the relationship between later time points of CF and EBP yielded nonsignificant findings, with no significant direct, indirect or, total effects being found. First, fall of first grade CF (CF3; β = -.04; t = -1.45; p = 0.154) and fall of second grade CF (CF5; β = -.00; t = -0.01; p = 0.988) failed to predict EBP to a statistically significant degree. Furthermore, both spring of kindergarten CF (CF2; β = -.02; t = -1.10; p = 0.277) and spring of first grade CF (CF4; β = .04; t = 1.27; p = 0.213) were found to predict EBP no better than chance levels. Finally, the correlation between spring of second grade CF (CF6) and EBP was nonsignificant (r = -.034; p = 0.110).

Hypothesis 3. The current study hypothesized that CF2 would significantly mediate the relationship between WM1 and EBP as measured in the spring of second grade (See Figure 6). Mediation can be determined through analysis of the relationships between the independent variable, mediating variable, and dependent variable. First, the independent variable must have a significant direct effect on the dependent variable when the mediating variable is not present. This can be achieved through a separate regression analysis; however, this was not done as part of the current study as the other conditions for mediation were not present. Next, a significant direct effect of the independent variable on the mediating variable must be present (i.e., WM on CF). Finally, and most importantly, the mediating variable must have a significant direct effect on the dependent variable (i.e., CF on EBP). If all conditions are present, and the mediating variable demonstrates a significant direct effect on the dependent variable, then a Sobel test can be conducted in order to determine if the mediating relationship is statistically significant. For the current hypothesis, the paths between WM1 and CF2 and the path between CF2 and EBP are of interest. First, the path from WM1 to CF2 was significant at the $p < .001$ level ($\beta = .20$; $t = 4.14$) providing support that early WM has a direct effect on later CF. Next, the path between CF2 and EBP did not reach significance suggesting no direct effect between early CF and later EBP ($\beta = -.02$; $t = -1.10$; $p = .277$). Given the lack of a direct effect of CF2 on EBP, the null hypothesis cannot be rejected as the model did not yield data supporting a direct effect of the hypothesized mediating variable on the dependent variable of interest. Analysis of subsequent timepoints provide the same pattern, with all timepoints of CF after fall of kindergarten failing to have a significant direct effect on EBP. Therefore, the current study failed to provide compelling evidence of a significant mediating relationship between CF and WM on EBP.

Hypothesis 4. In the present study, it was hypothesized that WM1 would significantly predict CF2 (See Figure 6). As already mentioned, the path from WM1 to CF2 yielded a moderate effect that was significant at the $p < .001$ level ($\beta = .20$; $t = 4.14$), providing robust evidence to reject the null hypothesis. It can be concluded that WM1 had a measurable and statistically significant impact on CF2. In fact, for every standard deviation unit increase in

WM1, participants' CF2 scores increased by .2 standard deviation units. In addition, the influence of WM1 on subsequent time points of CF was also analyzed to determine how far in advance early WM could predict later CF. The paths leading from WM1 to CF3 ($\beta = .06$; $t = 2.24$; $p = 0.031$), CF4 ($\beta = .13$; $t = 2.60$; $p = .013$), and CF5 ($\beta = .04$; $t = 2.12$; $p = 0.041$) all yielded significant direct effects. No significant direct effect was found between WM1 and CF6 ($\beta = .04$; $t = 1.65$; $p = 0.11$).

Similar to the relationship observed between WM1 and a majority of later timepoints for CF, WM2 also had significant direct effects on CF3 ($\beta = .18$; $t = 5.49$; $p < .001$) and CF5 ($\beta = .14$; $t = 5.79$; $p < .001$). As for the relationship between WM2 and both CF4 ($\beta = .05$; $t = 1.32$; $p = .194$) and CF6 ($\beta = -.00$; $t = -0.03$; $p = .980$), only significant indirect effects were noted (please see Table 5). In regard to WM3, direct effects on CF4 ($\beta = .08$; $t = 2.80$; $p = .008$) and CF5 ($\beta = .08$; $t = 2.61$; $p = .013$) were found; however, no significant direct effect was found on CF6 ($\beta = .03$; $t = 0.91$; $p = .370$). In further support of WM's influence on CF, WM4 significantly directly influenced CF5 to a small degree ($\beta = .08$; $t = 3.11$; $p = .003$), and CF6 to a small degree ($\beta = .08$; $t = 4.24$; $p < .001$); while WM5 significantly directly influenced CF6 scores to a small degree, as well ($\beta = .09$; $t = 3.48$; $p = 0.001$). Finally, analysis of indirect effects indicated numerous significant results; however, given that the indirect effects of WM on CF were entirely through later time points of WM (e.g., WM2), please refer to Table 5 for a summary. In sum, prior timepoints of WM influenced later development of CF in both direct and indirect ways. In particular, there was evidence supporting a direct influence of WM on CF, as well as an indirect influence primarily through later timepoints of WM.

Hypothesis 5. The current study hypothesized that fall of 2nd grade measures of WM and CF would have a stronger relationship to spring of 2nd grade EBP when compared to earlier time points of WM and CF (See Figure 5). When using path analysis, review of β's between different variables allows researchers to compare the relative magnitude these variables have on a given dependent variable. In the current study, β coefficients from each time point of WM and CF were visually inspected in order to determine if their strength increased as the measure of WM or CF grew temporally closer to the spring measure of EBP. As already mentioned, fall of kindergarten CF and fall of second grade WM had the only statistically significant direct effects on EBP, with low scores of each measure of EF predicting higher scores of EBP. Furthermore, relative to the earlier time points of WM (i.e., WM1, WM2, WM3, WM4), WM5 had the highest β value ($\beta = -.08$; $t = -3.52$; $p = .001$) indicating it was the strongest related time point of WM to EBP. In contrast, of all the time points for CF, CF1 showed the strongest direct effect on EBP ($\beta = -.06$; t = -2.03; p = 0.049), thus failing to provide evidence in support of the hypothesis that the strongest relationship between EF and EBP would be between measures captured closer in time to the measure of EBP.

Findings Related to Standardized Covariances

In the present study, numerous correlations were specified among WM and CF measured at the same point in time. More specifically, a correlation was hypothesized to exist between WM and CF for each time point (See Figure 5). This, rather than a directional influence, was hypothesized to exist because no time precedence could be specified for measures of EF made at the same point in time. As such, path analysis allowed for this relationship to be observed through the lens of a standardized covariance, or correlation, rather than through the lens of a Beta value, as is done with regression.

First, all timepoints of WM and CF were significantly related to one another, with exception to the spring of 1^{st} grade scores (WM4 and CF4; $r = .13$; p = .058). One potential explanation for the lack of a significant relationship being found during the spring of 1st grade may be due to a skewed distribution found in CF during the spring of 1st grade due to many of the participants scoring very well (mean = 16.33; SE = .09). In fact, this skewed distribution within CF was part of the decision to change the administration of the DCCS to a computerized format that scored for accuracy and RT, beginning in the fall of second grade. Despite not observing a significant relationship between CF4 and WM4, the relationship did approach significant levels ($p = .058$), and was in the positive direction as was originally hypothesized.

Following the change in administration of the DCCS during the fall of second grade, the relationship between WM and CF reached significant levels yet again (WM5 and CF5; $r = .1$; $p < .001$). Thus, creating more variability among the scores on the DCCS, by instituting a RT component into the scoring, improved the significance of the correlation between WM and CF relative to the spring of first grade where many of the participants had mastered the DCCS. Finally, the strongest relationship between WM and CF existed during the fall of kindergarten, as evidenced by a weak correlation ($r = .28$; $p < .001$), with the weakest significant relationship being found between WM and CF during the spring of second grade (WM6 and CF6; $r = .07$; $p = .025$). Interestingly, the strength of the correlation generally weakened as the participants aged.

Working Memory's Contribution to Cognitive Flexibility

The current study sought out to investigate the degree to which WM contributes to the unfolding of CF, with the underlying hypothesis that higher levels of WM would contribute to higher levels of CF. Convergent with this idea, findings from the path analysis are indicative of a consistent pattern, regardless of the grade level of the participants. In particular, the predictive utility of WM on CF was found across multiple time points. As already mentioned, WM1 was found to significantly predict CF2 scores during kindergarten ($\beta = .20$; $t = 4.14$; $p < .001$).

Additionally, WM1's predictive strength for CF was found to span multiple grade levels as a significant direct effect was noted on CF3 (β = .06; t = 2.24; p = .031), CF4 (β = .13; t = 2.60; p = .013), and CF5 (β = .04; t = 2.12; p = .041).

In addition to finding predictive strength from kindergarten to second grade, a notable pattern emerged within each grade level. In particular, the significant relation between WM1 and CF2 was already noted for kindergarten. Similarly, WM3 was found to significantly predict CF4 scores during first grade (β = .08; t = 2.80; p = .008), and WM5 was found to significantly predict CF6 scores during second grade (β = .09; t = 3.48; p = .001). Thus, a child's WM during the fall of each grade level significantly predicted their CF during the spring of that same year with a small to moderate strength. Taken together with the previously mentioned finding that the predictive strength of WM on CF spanned multiple grade levels (e.g., WM1 significantly predicted CF5), the current results provide powerful evidence in support of the theory that WM is important to the development of CF.

Chapter V: Discussion

The overall goal of this study was to further uncover influence that two major components of EF (i.e., WM and CF) have on EBP, while simultaneously analyzing the development of these components during the early elementary years while controlling for sex, race, and SES. In regard to the former, it was hypothesized that lower levels of WM and CF in kindergarten would be predictive of higher EBP by the end of second grade. First and foremost, the current study succeeded in providing consistent evidence with prior work investigating the complex role that EF has in EBP (e.g., Antshel et al., 2014; Schoemaker et al., 2013). In particular, WM in kindergarten was not predictive of later EBP in second grade; however, WM, as measured in the fall of second grade, did yield a small, yet statistically significant relationship with the spring of second grade measure of EBP. Relative to CF, WM appeared to demonstrate a slightly stronger relationship to EBP but this difference may not be meaningful. As for CF, evidence supporting the original research hypothesis was found, indicating that fall of kindergarten WM was a significant predictor of later EBP in spring of second grade. In fact, for every unit increase in CF scores, children exhibited -.06 standard deviations fewer symptoms of EBP per teacher reports. Thus, it appears that rigid thinking, or an inability to shift between different mental sets when solving a problem, may contribute to a higher probability of later EBP possibly due to the child adopting an inflexible perspective when making behavioral choices. Taken together, the current study provides evidence that early WM and CF does play somewhat of a role in the later manifestation of EBP.

In addition to investigating the role of EF in EBP, the current study also sought out to investigate the degree to which EF demonstrates a hierarchical development in which WM contributes to the unfolding of CF (Best et al., 2009; Best & Miller, 2012; Davidson et al., 2006;

Garon et al., 2008). Based on the results, earlier time points of WM were found to significantly influence later time points of CF, indicating that successful WM early on leads to successful development of CF at later time points. Furthermore, WM's influence on CF was found to span a two-year gap, with a small but meaningful direct effect found between fall of kindergarten WM and fall of second grade CF. When trying to explain this effect, it seems clear that being able to successfully hold information in present awareness is a necessary prerequisite for being able to shift between different rulesets for solving problems. In other words, an individual must be able retain a ruleset before being able to flexibly shift between two or more rulesets – and now the present study provides evidence to support this claim.

Although it was not one of the current study's main goals, results were supportive of prior research that has demonstrated a protracted development of WM and CF (e.g., Best & Miller, 2010; Best et al., 2009). In particular, visual inspection of mean scores associated with each component of EF show a general linear growth pattern, with improvements being noted in each component from timepoint to timepoint. Unfortunately, given the change in administration in the DCCS during the fall of second grade wave of data collection, growth cannot be determined given the score was on a different scale. In spite of this data complication, earlier timepoints of CF did exhibit significant predictive relationships with the alternative-scored version of the DCCS, demonstrating that despite the scoring change, performances on each version were related to one another. Thus, including reaction time as well as accuracy does not fundamentally change the construct being measured, as some accuracy-only time points (e.g., CF1, CF2, CF3, CF4) still significantly predicted accuracy and reaction-time measures of WM (e.g., CF5 and CF6).

In terms of the relationship between WM and CF, the current study provides evidence in support of Miyake et al.'s (2000) assertion that different EF tasks typically manifest low correlations between each other. The correlations between WM and CF ranged from nonsignificant to weak, with the strongest being the association between both fall of kindergarten measures. Thus, despite having theoretical connections to one another, it appears individual tasks measuring WM and CF have weak relations during the early elementary years; suggesting that the executive functions these tasks tap into exist as discrete, unrelated cognitive functions in children, or that our measurement of these executive functions must improve in order to better capture what is common among both of these important cognitive abilities.

Implications of Findings

Development of executive functioning. As already stated, the current study contributes to previous research in supporting the claim that WM and CF show developmental growth during the early elementary years (e.g., Best & Miller, 2010; Best et al., 2009; Diamond, 2006). From increases in a child's ability to hold more information in present awareness, to increases in their ability to flexibly solve problems with different mental sets, it is clear that children undergo growth in these key components of EF from kindergarten to second grade, and that earlier levels of each component predict later performance. Moreover, the current study provides clear evidence that EF manifests with unique, but related components during the early elementary years, providing increasing support for the *Unity/Diversity Framework* put forth by Miyake et al. (2000) despite not using latent-variable analysis. In particular, as the cohort aged from each wave of data collection to the other, the relationship between WM and CF continued to differentiate as evidenced by a decreasing strength in the correlation between the two. Despite this conclusion, it is important to note that the current model's fit was unable to be tested formally due to the use of complex survey data.

Most interestingly, the current findings show consistency with theoretical work aimed at explaining the differential growth observed in WM and CF (e.g., Davidson et al., 2006), providing researchers with one of the first studies that quantified the degree to which WM contributes to CF development using path analysis. In particular, the current evidence supports the idea that CF develops as a response to development of WM, in that successful CF must require WM to a certain degree as the individual must be able to hold information in their present awareness in order to flexibly shift between differing mental sets for solving a given problem. Depending on the temporal distance between the measure of WM and CF, the strength of the relationship ranged from nonsignificant to moderate, with fall of kindergarten WM showing the largest and strongest predictive relationship to CF as measured in the spring of kindergarten. In fact, WM1 showed a stronger contribution to CF2 than CF1 scores, suggesting that development of WM is more important to later manifestations of CF than earlier levels of CF. Not all time points of CF demonstrated this pattern where early WM was a stronger predictor of CF than earlier time points of CF; however, the findings are still important to our understanding of the development of CF and WM. More rigorous research is necessary to better understand this complex relationship, especially at the latent-variable level.

EF's Relationship to Externalizing Behavior Problems. The results of this study indicate that of the two major components of EF captured within the ECLS-K dataset, WM had the larger impact on subsequent development of EBP relative to CF, after controlling for sex, race, and SES. More specifically, WM5 was the one of two statistically significant predictors, with a small negative relationship being found indicating that children with high WM exhibited lower amounts of EBP. This small effect can probably be explained by the child's ability to hold more information in their present awareness, better allowing them to follow rules, directions, and remember the norms for a given situation. For example, a child with higher WM would better be able to direct their behavior in accordance with rules just delivered by their classroom teacher whereas the child with lower WM would struggle to remember rules and make behavioral choices in line with the current situation's demands.

When compared to prior work, the current study mirrors findings from meta-analytic work conducted with preschoolers by Schoemaker et al. (2013) that also found a small effect size for WM's contribution to EBP. Taken together, it appears that the WM's impact on EBP remains stable from the preschool years through second grade; however, this is only speculation as the current study only captured EBP during the spring of second grade. Moreover, no other time points of WM reached significant levels, indicating that temporal proximity, or how close the measure of WM is to the measure of EBP in time, may be important to observing the relationship. Future work should utilize the ECLS-K dataset to uncover the relationship between EF and EBP measured in kindergarten through second grade so more robust conclusions can be drawn.

Notably, given that only a small effect was found between WM5 and EBP, it appears that the current study failed to provide strong enough evidence to suggest a change in best practices for dealing with EBP. In other words, the resources necessary to assess second grade WM on a wide scale during the fall in efforts to identify children for intervention may not be justified if only a small reduction in EBP may potentially be observed. This does not rule out the utility of assessing WM and directing accommodations or intervention efforts for children demonstrating significant levels of EBP with comorbid WM deficits. Especially when known future consequences (e.g., juvenile delinquency, crime, substance abuse, employment and relationship difficulties, and violence; Liu, 2004; McKee, Colletti, Rakow, Jones, & Forehand, 2008) of a pattern of early EBP are taken into account. More research is necessary before resource-intensive decisions are made on a systems-wide level. In the present, it appears early WM is important to EBP to a small degree; however, there are a multitude of other variables (e.g., prior achievement, SES, and preschool environment) that may prove to be a better focus for assessment and intervention efforts aimed at reducing EBP.

In contrast to the statically significant direct effect of WM (i.e., WM5) on EBP, all time points of CF failed to reach significance with exception to the fall of kindergarten measure (i.e., CF1). Notably, CF1 showed the strongest relationship to EBP, with a small negative relationship being observed; however, the direct effects on EBP generally weakened after that time point. CF demonstrated the hypothesized relationship in that children with higher levels of CF during the fall of kindergarten demonstrated a smaller magnitude EBP symptoms as measured in the spring of second grade. Thus, it appears that early CF may influence the later onset of EBP, but exactly how is still unclear. Potential explanations range from CF having a direct impact on the early onset of EBP to CF's influence on other important variables that may mediate the later development of EBP.

The current study's finding that early CF significantly predicted later EBP provides replication to the Schoemaker et al. (2013) meta-analysis conducted with preschoolers that also found a small effect size of CF on concurrent EBP. Thus, it appears that CF, as measured at school entry, is a significant predictor of EBP during the spring of second grade; however, it appears that CF explains less and less of the variance in EBP as children age. More research investigating the relationship between CF and concurrent EBP is necessary before strong conclusions are drawn.

One explanation for the lack of significant findings for later timepoints of CF could be that CF may have more of an impact on prosocial behaviors, as opposed to the successful inhibition of EBP. In other words, maybe a child's CF does not prevent the onset of EBP so much as it fosters prosocial behaviors through the child's ability to use multiple perspectives when making behavioral decisions. Prior research has demonstrated a relationship between CF and socially cooperative behaviors at seven, nine, and 11 years old (e.g., Ciairano et al., 2006) through a situation where children were expected to work on a puzzle together. In addition, general EF has been linked to the development of theory of mind, or the ability to take the perspective and beliefs of another human (e.g., Bock et al., 2015; Hughes, 1998). This relationship of EF to theory of mind has natural connections to our current theoretical understanding of CF, as well as implications for our understanding of the social deficits that characterize individuals with Autism Spectrum Disorder (ASD). Regarding the latter, research supports the notion that individuals with ASD generally have EF deficits (Johnson, 2012), which may explain at least some of the social deficits that characterize the disorder. Finally, prior research using children aged seven to 12 years demonstrated a stronger contribution of CF on social understanding above and beyond what was explained by WM, Inhibition, age, and vocabulary (Bock et al., 2015). Thus, CF may serve more as a proactive variable in fostering adaptive social traits, as opposed to helping the child with EBP to inhibit impulsive decisions, especially as the child ages.

School-Based Implications

The current study sought to better understand the development of EF and its relationship to EBP in order to equip school-based practitioners with another tool for helping children who are presenting with problematic behaviors in the school environment. Despite having the intention of demonstrating a medium to strong effect of WM and CF on EBP, current findings yielded only two significant timepoints of EF (i.e., small effect of CF1, small effect of WM5) for predicting EBP during the spring of second grade. Of all the variables used to predict later EBP, the only other significant predictors included key demographic and background variables that were controlled for (i.e., sex, race, and SES). In spite of not finding significant results across the board, the current study does provide some meaningful information pertaining to EF and how a student's demographic information may put them at risk for future EBP. With this knowledge, school-based practitioners can organize systems-level services for students identified at risk, as prevention can be a powerful tool for ensuring that a pattern of EBP does not unfold later in the child's life.

As already mentioned, participants identifying as black or Hispanic were found to demonstrate significantly different levels of EBP relative to white participants within the current study; with black participants demonstrating significantly more EBP and Hispanic participants demonstrating significantly less. In regard to black participants demonstrating more EBP than white comparisons, a significant negative correlation was also noted between being black and SES, highlighting a potential relationship that needs to be investigated further. More specifically, is there an interaction between SES and another variable important to the later onset of EBP that impacts children identifying as black differently than other racial groups? Taking into account the small effect that also was found between SES and EBP, it appears the current study provides evidence that children coming from lower SES environments have a higher risk for demonstrating increased amounts of EBP.

When compared to prior meta-analytic work (e.g., Korous, Causadias, Bradley, & Luthar, 2018), the impact of SES on racial minorities and EBP comes to no surprise as substantial evidence already exists linking impoverished conditions to a greater risk for internalizing and EBP. From observations of increased parent-child conflict (e.g., Wolfe & Renk, 2017) to emerging evidence linking low-SES environments with increased sleep problems within children (e.g., Lam & Chung, 2017), it is clear that low-SES environments put children at-risk for the development of later behavior problems. With this understanding of the literature in place, the onus is placed on schools, especially those serving low-SES communities, to prepare their teachers and other personnel for implementing culturally-sensitive practices. In addition, to the extent that it is possible, school-home collaboration should be fostered so parent training can be offered to families in need.

In addition to the observed relationship between race, SES and EBP, the current study provides consistent results with prior work that has already established a link between race, SES, and EF. In particular, Little (2017) demonstrated that Black students were found to exhibit significantly lower WM and CF relative to White students at kindergarten entry. The current study demonstrates the same finding, in the order of a small to medium effect of being Black and exhibiting significantly lower levels of WM and CF during the fall of kindergarten. Although not the focus of the current study, the importance of addressing racial discrepancies in EF at school-entry appears to be rising as mounting evidence emerges demonstrating that these discrepancies exist; especially when taking into account the growing literature base demonstrating a link between low EF and academic underachievement, as well as social and behavioral troubles (e.g., Best et al., 2011; Hughes & Ensor, 2011; Morgan et al., 2017; Sulik et al., 2015). If minority students are coming in with lower EF and higher levels of EBP, then efforts to reduce these discrepancies must occur on a school and societal level. From increasing family-school collaboration to improving culturally-informed practices on a curricular and social-emotional level, schools must seek to help families no later than kindergarten as we now see the discrepancies already exist as early as then. From offering preventative services and improving connections with community services to implementing interventions with children demonstrating weaknesses early on, our schools are community hotspots for providing children the help they need.

Limitations

One issue that the current study suffered from, similarly to many other studies of EF, is the issue of task impurity (Miyake et al., 2000). Task impurity is defined as the contamination of multiple executive functions within a single measure of EF, making it difficult to parse out the individual influence specific components have on another variable. Even broader, the measurement of EF seems to be a controversial issue, with some researchers electing to use individual tests that purport to measure a component of EF directly (e.g., Ciairano et al., 2006, Miyake et al., 2000; Zelazo, 2006), while other researchers have seemingly discredited the use of these tests in favor of rating scales that rely on the subjective interpretation of informants or self-report (e.g., Barkley, 2014).

Within the current study, the ECLS-K dataset relied on two individually-administered tests of EF (i.e., Numbers Reversed for WM, DCCS for CF), which on the surface, do not have easily-made connections to the theoretical implications the broad constructs of WM and CF have on behavior. To be candid, having a child successfully recite auditorily-presented strands of numbers backwards, or be able to categorize cards with pictures of animals and boats, does not have clear connections to the social situations in which the constructs these tasks measure should influence. Moreover, the logical implications of a cognitive ability for directing attention in an effortful manner and holding information in present awareness are easy to identify for countless social situations that children and adults come across on a daily basis. Being able to flexibly shift between different perspectives is surely a necessary skill for the social arena, as empathic behavior is seemingly built upon this foundational skill.

Despite the theoretical connections these constructs have to the successful management of behavior, the current study failed to produce significant effects to the degree that was predicted, especially with regard to WM and CF’s influence on EBP. As such, the fault may lie in the way the constructs were measured. It is possible that using rating scale measures of WM and CF, such as that offered on the BRIEF (Gioia et al., 2000) or Comprehensive Executive Function Inventory (CEFI; Naglieri & Goldstein, 2013), may have offered improved predictive accuracy for EF on EBP, as the way WM and CF is measured with these tools is fundamentally different than what is offered on direct tests of WM and CF such as the Numbers Reversed Test of the WJ-III-COG (Woodcock et al., 2001b) or the DCCS (Zelazo, 2006).

Another limitation of the current study again relates to the measurement of EF. As already stated, Miyake et al. (2000) demonstrated improved correlations between components of EF (WM, CF, and Inhibition) through use of a latent-variable procedure. The current study was unable to use this form of SEM due to the ECLS-K dataset only containing one indicator, or one measure, of each component of EF. Ideally, future research would be able to administer multiple measures of each component, extract what is common among the measures, and use this latent variable as the predicting influence on EBP. As the current study was unable to do this, path analysis was the correct statistical technique to use; however, the role of task impurity in attenuating the findings cannot be ruled out.

Finally, as with all forms of regression, a significant limitation that exists involves the selection of variables that are included in the model that is eventually estimated. As Keith (2014) clearly puts in his text on multiple regression and SEM, regression coefficients have the potential to drastically change by the inclusion of additional variables, or lack thereof, for any model. When formulating the model that is eventually tested, one must take into account theory, logic, and prior research in order to inform the variables that are included so the resulting coefficients are as accurate as possible. As such, the ideal model would include all of the influencing variables and common causes of your independent and dependent variables that exist, so that the resulting regression coefficients are accurately pulled out with consideration of each variable within the model. Unfortunately, in practice, this is difficult to implement as theory is constantly advancing, as well as our understanding of what the important variables are for a given group of variables. When using an extant dataset, this difficulty is further compounded as the researcher no longer has control over which variables are measured at the time of data collection.

In conclusion, like most research, the current study exhibited some methodological weaknesses that likely influenced the presenting results in important ways. Issues such as the measurement of EF and the inclusion of other important covariates were discussed. Despite these weaknesses, the use of a large weighted sample that represented a nationally-representative cohort of 2010-2011 kindergarten students on important background variables such as race and SES, and the use of path analysis to understand the relative influence of key variables in the discussion of EF and EBP, contributes strong evidence supporting an important role of EF in EBP.

Future Research

Given the limitations of using an extant dataset, the current study did not include other known important variables to EF and EBP within its model. As such, future research should make efforts to include other important correlates, so the relative influence of each variable on EF and EBP can be better understood. For example, research investigating the role of activity settings in children's early learning behaviors within full-day kindergarten classrooms yielded interesting findings (Ansari & Purtell, 2017). In particular, using the same sample as the current study, Ansari and Purtell (2017) found that the way kindergarten teachers structure their classroom had a significant impact on children's early learning as well as their developing EF. Children coming from classrooms that spent a larger proportion of time in teacher-directed whole-group instruction demonstrated the largest gains in academic achievement relative to other activity-setting types, such as teacher-directed small-group dominated classrooms, as well as child-selected activity dominant classrooms. Interestingly, the opposite was true for the development of EF, with children who spent a majority of their time in classrooms that used child-selected activities demonstrating the largest gains in development of their CF. It is important to note that no significant differences were found in children's social-emotional functioning, which was measured using the SSRS composite of social skills, internalizing behavior problems, and EBP. Despite this finding, the authors conclude that other classroom variables could have an impact on social-emotional functioning, such as the emotional climate set by the teacher. Thus, future work must be done to further entangle the role that teaching styles have on developing EF, as well as on the presence of EBP.

In addition to understanding the role of teaching in EF and EBP, more research is needed to parse out the relative influence that WM, CF, and Inhibition have on important outcomes such as academic achievement and social-emotional functioning. Future research should make all efforts to include all three factors of EF when estimating the relative influence that these variables have on any given dependent variable; especially when the numerous CFA studies that have demonstrated a clear best fit for the three-factor model are taken into account (e.g., Friedman et al., 20008; Miyake et al., 2000; Rose et al., 2011); in addition to our understanding of how failing to include an important and related variable within a regression-based model can drastically change the resulting coefficients.

Other important factors that should be accounted for in future research are easily identified. A review conducted by Campbell, Shaw, and Gilliom (2000) identified numerous variables that put children at risk for the later development of EBP. These variables include early levels of hyperactivity and aggression, high levels of negative parenting and familial stress, and other sociodemographic and neighborhood influences that also likely have an impact on a child's ability to manage their behaviors in the school environment. As already mentioned, parenting practices (e.g., coercive parenting vs. sensitive parenting) have been found to affect a child's EF, as well as the degree to which children exhibit EBP (e.g., Smith, Dishion, Shaw, Wilson, Winter, & Patterson, 2014; Sulik et al., 2015).

Further complicating these intricate relationships is the bidirectional relationship that has been observed between parenting practices and EBP. For example, children who demonstrate early levels of hyperactivity and aggression have not only demonstrated to be more at-risk for later EBP, but they also tend to be on the receiving end of more harsh parenting practices (e.g., Wootton, Frick, Shelton, & Silverthorn, 1997). In other words, the child exhibits a problematic behavior that frustrates the parent. This frustration manifests in an ineffective parenting response characterized by frustration and anger, leading the child to respond with equal or greater intensity. This cycle continues, inadvertently reinforcing the problematic behaviors, and leading the child to learn this negative pattern of relating to others (Smith et al., 2014). Finally, although the current study controlled for SES, prior research has demonstrated a clear link between children from low SES backgrounds and greater EBP during the preschool years (Huaqing & Kaiser, 2003). SES may be related to other important background variables such as a child's preschool environment, or lack thereof. Exposing a child to preschool may better prepare them for the onset of kindergarten, giving them a head start to the behavioral and academic expectations they will be expected to meet. In sum, the current study helps advance the fields of EF and EBP; however, important methodological weaknesses must be accounted for in future work in order to continue advancing our understanding of EF and EBP.

Final Remarks

The current study set out to better understand the intricate relationship between EF and EBP, while simultaneously analyzing the degree to which WM contributes to the unfolding of later CF during the early elementary years, while controlling for race, sex, and SES. Two timepoints of EF proved to produce significant, small effects. In particular, fall of kindergarten CF yielded a negative relationship with spring of second grade EBP, indicating that a child's CF at school entry predicted with a negative relationship the degree to which the same child's second-grade teacher endorsed symptoms of EBP later down the road. The same negative relationship was also observed between WM, as measured in the fall of second grade, and EBP, but to a slightly stronger degree than CF. As for the development of CF, the current study provides rigorous support to the notion that WM is important to the eventual development of CF, with predictive strength being observed between WM and CF spanning two grade levels. Similarly, fall WM in each grade level was found to predict spring CF, thus demonstrating that the manifestation of CF is likely dependent on elements of WM.

Taken together, these results have particular value due to the sample that was used, as well as the statistical methodology. For example, the current study utilized a nationally-representative sample of kindergarteners from the 2010-2011 cohort. Despite losing a large portion of the sample to missing data and sample attrition, the use of weight variables still allowed for generalization to other kindergarteners who were entering school during the same school year. The degree to which these kindergarteners are similar to more current kindergarteners has yet to be determined. As for the methodology, the current study utilized two well-respected measures of WM and CF, as well as a powerful statistical technique for determining the relative magnitude of influence that independent variables have on dependent variables. Thus, the current study offers some of the first evidence estimating the magnitude of influence that WM has on CF development from kindergarten to second grade. Finally, current results add to a growing literature base that demonstrates the importance of early EF to later EBP during the elementary years.

In closing, school-based practitioners can utilize the current study as a preliminary look at the efficacy of using measures of WM and CF to predict later EBP. Although significant findings were found, more research is necessary before assessments of EF could be relied upon as a tool for informing prevention measures against later EBP in children. Despite the current measures of WM and CF not demonstrating strong predictive power, school-based practitioners may begin to conceptualize students struggling with controlling their behavior through the lens of EF. Similarly, as the research advances, concrete guidance for optimal parenting strategies to foster EF may emerge as well. Understanding how weaknesses in EF may impact a child may prove to be a beneficial route for ameliorating the difficulties experienced by children with EBP.

References

Acock, A.C. (2013). *Discovering Structural Equation Modeling Using Stata.* College Station, TX: Stata press.

Achenbach, T., & Edelbrock, C. (1983). *Manual for the child behavior checklist and revised child behavior profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M., & Rescorla, L.A. (2001). Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (DSM-5). Arlington, VA: American Psychiatric Publishing.

Ansari, A., & Purtell, K.M. (2017). Activity settings in full-day kindergarten classrooms and children's early learning. *Early Childhood Research Quarterly, 38*, pp. 23-32. doi: 10.1016/j.ecresq.2016.09.003.

Antshel, K. M., Hier, B. O., & Barkley, R. A. (2014). Executive functioning theory and ADHD. In S. Goldstein, J. A. Naglieri, S. Goldstein, J. A. Naglieri (Eds.), *Handbook of executive functioning* (pp. 107-120). New York, NY, US: Springer Science + Business Media. doi: 10.1007/978-1-4614-8106-5_7.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), pp. 91-93. Doi: 10.1016/j.tjem.2018.08.001.

Barkley, R.A. (2014). *Attention-deficit hyperactivity disorder: Fourth edition.* New York, NY: The Guilford Press.

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, *36*(2), pp. 137-161. Doi:10.1080/87565641.2010.549877.

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

Best, J.R., Miller, P.H., & Jones, L.L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review, 29*(3), pp. 180-200. Doi: 10.1016/j.dr.2009.05.002.

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*, pp. 327-336. Doi: 10.1016/j.lindif.2011.01.007.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T., ... & Johnson, M. A. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry*, *159*(1), pp. 36-42. Doi: 10.1176/appi.ajp.159.1.36.

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., . . . & Faraone, S. V. (2004). Impact of Executive Function Deficits and Attention-Deficit/Hyperactivity Disorder (ADHD) on Academic Outcomes in Children. *Journal of Consulting and Clinical Psychology*, *72*(5), pp. 757-766. Doi: 10.1037/0022-006X.72.5.757.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*, pp. 109-160. Doi: 10.1017/S0140525X06009034.

Blakey, E., Visser, I., & Carroll, D.J. (2016). Different executive functions support different kinds of cognitive flexibility: Evidence from 2-, 3-, and 4-year-olds. *Child Development, 87*(2), pp. 513-526. Doi: 10.1111/cdev.12468.

Bock, A.M., Gallaway, K.C., & Hund, A.M. (2015). Specifying links between executive functioning and theory of mind during middle childhood: Cognitive flexibility predicts social understanding. *Journal of Cognition and Development, 16*(3), pp. 509-521. doi: 10.1080/15248372.2014.888350.

Buckhalt, J.A., El-Sheikh, M., & Keller, P. (2007). Children's sleep and cognitive functioning: Race and socioeconomic status as moderators of effects. *Child Development, 78*(1), pp. 213-231. doi: 10.1111/j.1467-8624.2007.00993.x.

Campbell, S.B., Shaw, D.S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology, 12*(3), pp. 467-488. doi: 10.1017/S0954579400003114.

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, *37*(5), pp. 715-730. doi:10.1037/0012-1649.37.5.715

Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: Effect of cue transparency on children's cognitive flexibility. *Developmental psychology*, *45*(3), pp. 782-797. doi: 10.1037/a0015409.

Ciairano, S., Bonino, S., & Miceli, R. (2006). Cognitive flexibility and social competence from childhood to early adolescence. *Cognition, Brain, & Behavior*, *X*(3), pp. 343-366. Retrieved from http://www.ascred.ro/.

Conklin, H. M., Luciana, M., Hooper, C. J., & Yarger, R. S. (2007). Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. *Developmental neuropsychology*, *31*(1), pp. 103-128. doi: 10.1207/s15326942dn3101_6.

Cottrell, J. M., Newman, D. A., & Roisman, G. I. (2015). Explaining the black–white gap in cognitive test scores: Toward a theory of adverse impact. *Journal of Applied Psychology*, *100*(6), pp. 1713-1736. doi:10.1037/apl0000020

Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *The Quarterly Journal of Experimental Psychology*, *65*(2), pp. 209-232. doi: 10.1080/17470210903204618.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), pp. 2037-2078. doi: 10.1016/j.neuropsychologia.2006.02.006.

Deater-Deckard, K., Dodge, K.A., Bates, J.E., & Pettit, G.S. (1996). Physical discipline among African American and European American mothers: Links to children's externalizing behaviors. *Developmental Psychology, 32*(6), pp. 1065-1072. doi: 10.1037/0012-1649.32.6.1065.

Deater-Deckard, K., Dodge, K.A., Bates, J.E., & Pettit, G.S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Developmental Psychology, 10*(3), pp. 469-493. Available from: http://www.apa.org/pubs/journals/dev/.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan executive function system.* San Antonio, TX: Pearson (The Psychological Corporation).

Diamond, A. (2006). The early development of executive functions. In E. Bialystock & F. I. M. Craik (Eds.), *The early development of executive functions. Lifespan cognition: Mechanisms of change* (pp. 70-95). Oxford, UK: Oxford University Press.

Diamond, A., Carlson, S.M., & Beck, D.M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology, 28*(2), pp. 689-729. doi: 10.1207/s15326942dn2802_7.

Dodge, K.A., Pettit, G.S., & Bates, J.E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65,* pp. 649-665. doi: 10.1111/j.1467-8624.1994.tb00774.x.

Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia, 33*(3), pp. 261-268. Retrieved from https://www.journals.elsevier.com/neuropsychologia.

Dunn, L.M., & Dunn, D.M. (2007). *Peabody picture vocabulary test* (4th ed.). San Antonio, TX: Pearson.

Friedman, N.P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86,* pp. 186-204. doi: 10.1016/j.cortex.2016.04.023.

Friedman, N.P., Miyake, A., Corley, R.P., Young, S.E., DeFries, J.C., & Hewitt, J.K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), pp. 172-179. Retrieved from http://www.jstor.org/stable/40064389.

Friedman, N.P., Miyake, A., Robinson, J.L., & Hewitt, J.K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*(5), pp. 1410-1430. doi: 10.1037/a0023750.

Friedman, N.P., Miyake, A., Young, S.E., DeFries, J.C., Corley, R.P., & Hewitt, J.K. (2008). Individual differences in executive functions are almost entirely genetic. *Journal of Experimental Psychology: General*, *137*(2), pp. 201-225. doi: 10.1037/0096-3445.137.2.201.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), pp. 31-60. doi:10.1037/0033-2909.134.1.31.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), pp. 177-190. doi: 10.1037/0012-1649.40.2.177.

Gioia, G. A. Isquith. PK, Guy, SC, & Kenworthy, L. (2000). *Behavior rating inventory of executive function professional manual*. Odessa, FL: Psychological Assessment Resources.

Goldstein, S., Naglieri, J.A., Princiotta, D., & Otero, T.M. (2014). Introduction: a history of executive functioning as a theoretical and clinical construct. In *Handbook of executive functioning* (pp. 3-12). Springer New York. doi: 10.1007/978-1-4614-8106-5_1.

Gresham, F.M., & Elliott, S.N., (1990). *Social Skills Rating System.* Minneapolis, MN: Pearson Assessments.

Harter, S. (1985). *Manual for the self-perception profile for children*. University of Denver, Denver, CO.

Hinshaw, S.P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, *111*(1), pp. 127-155. doi: 10.1037/0033-2909.111.1.127.

Hongwanishkul, D., Happaney, K. R., Lee, W. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*(2), pp. 617-644. doi:10.1207/s15326942dn2802_4

Huaqing, C. & Kaiser, A.P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Topics in Early Childhood Special Education, 23*(4), pp. 188-216. doi: 10.1177/02711214030230040201.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British journal of developmental psychology*, *16*(2), pp. 233-253. doi: 10.1111/j.2044-835X.1998.tb00921.x.

Hughes, C., & Ensor, R. (2008). Does executive function matter for preschoolers' problem behaviors? *Journal of Abnormal Child Psychology, 36*(1), pp. 1-14. doi: 10.1007/s10802-007-9107-6.

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,* pp. 663-676. doi: 10.1016/j.jecp.2010.06.005.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), pp. 2017-2036. doi: 10.1016/j.neuropsychologia.2006.01.010.

Johnson, M. H. (2012). Executive function and developmental disorders: The flip side of the coin. *Trends in Cognitive Sciences*, *16*(9), pp. 454-457. doi:10.1016/j.tics.2012.07.00.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20,* 141-151. doi: 10.1177/001316446002000116

Kaufman, S.B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity. *Intelligence, 35*(3), pp. 211-223. doi: 10.1016/j.intell.2006.07.009.

Keith, T. Z. (2014). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling*. New York, NY: Routledge.

Korous, K.M., Causadias, J.M., Bradley, R.H., & Luthar, S.S. (2018). Unpacking the link between socioeconomic status and behavior problems: A second-order meta-analysis. *Development and Psychopathology*, *30*(5), pp. 1889-1906. doi: 10.1017/S0954579418001141.

Korsch, F., & Petermann, F. (2013). Agreement between parents and teachers on preschool children's behavior in a clinical sample with externalizing behavioral problems. *Journal of Child Psychiatry and Human Development, 45,* pp. 617-627. doi: 10.1007/s10578-013-0430-6.

Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 39*(1), pp. 47-63. Retrieved from https://www.wiley.com/WileyCDA/Brand/id-35.html.

Kwon, H., Reiss, A. L., & Menon, V. (2002). Neural basis of protracted developmental changes in visuo-spatial working memory. *Proceedings of the National Academy of Sciences*, *99*(20), pp. 13336-13341. doi: 10.1073/pnas.162486399.

Lam, C. B., & Chung, K. K. H. (2017). Associations of sleep problems with externalizing behaviors and preacademic performance: The moderating role of family socioeconomic status. *Infant and Child Development*, *26*(5), pp. 1-9. doi: 10.1002/icd.2026.

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology*, *49A*(1), pp. 29-50. Retrieved from http://journals.sagepub.com/.

Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. *Journal of Emotional and Behavioral Disorders*, *21*(3), pp. 211-224. doi: 10.1177/1063426611421007.

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school: Nationally representative evidence from the ECLS-K: 2011. *Educational Researcher, 42*(2), pp. 103-109. Doi: 10.3102/0013189X17698700.

Liu, J. (2004). Childhood externalizing behavior: Theory and implications. *Journal of Child and Adolescent Psychiatric Nursing, 17*(3), pp. 93-103. Doi:

Luciana, M., Conklin, H.M., Hooper, C.J., & Yarger, R.S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development, 76*(3), pp. 697-712. Retrieved from http://www.jstor.org/stable/3696460.

Mather, N., & Jaffe, L. E. (2016). *Woodcock-Johnson IV: Reports, recommendations, and strategies*. John Wiley & Sons.

McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of giants of psychometric intelligence research. *Intelligence, 37,* pp. 1-10. doi: 10.1016/j.intell.2008.08.004.

McGrew, K.S., & Woodcock, R.W. (2001). *Woodcock-Johnson III Technical Manual*. Itasca, IL: The Riverside Publishing Company.

Mckee, L., Colletti, C., Rakow, A., Jones, D.J., & Forehand, R. (2008). Parenting and child externalizing behaviors: Are the associations specific or diffuse? *Aggression and Violent Behavior, 13*(3), pp. 201-215. doi: 10.1016/j.avb.2008.03.005.

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), pp. 1373-1386. doi: 10.1111/j.1467-8624.2004.00746.x.

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

Miyake, A., & Friedman, N.P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), pp. 8-14. doi: 10.1177/0963721411429458.

Morgan, P.L., Hui, L., Farkas, G., Cook, M., Pun, W.H., & Hillemeier, M.M. (2017). Executive functioning deficits increase kindergarten children's risk for reading and mathematics difficulties in first grade. *Journal of Contemporary Education Psychology, 50*, pp. 23-32. doi: 10.1016/j.cedpsych.2016.01.004.

Naglieri, J.A. & Goldstein, S. (2013). *Comprehensive Executive Function Inventory Technical Manual.* Canada: Multi-Health Systems Inc.

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,* pp. 774-783. doi: 10.1016/j.ecresq.2013.07.005.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ogilvie, J., Stewart, A., Chan, R., & Shum, D. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology*, *49*(4), pp. 1063-1107. doi: 10.1111/j.1745-9125.2011.00252.x.

Raver, C. C., Blair, C., & Willoughby, M. (2013). Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. *Developmental psychology*, *49*(2), pp. 292-319. doi: 10.1037/a0028343.

Reynolds, C. R., & Kamphaus, R.W. (2015). *BASC 3: Behavior assessment system for children.* Bloomington, MN: PsychCorp.

Robert, M. & Savoie, N. (2006) Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology*, 18(3), pp. 378-397. Doi: 10.1080/09541440500234104.

Rose, S.A., Feldman, J.F., & Jankowski, J.J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*(5), pp. 1161-1175. doi: 10.1111/j.1467-7687.2011.01068.x.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer III, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*(2), pp. 297-330. doi: 10.111/j.1744-6570.2001.tb00094.x.

Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*(134), pp. 473-489. doi: 10.2307/2291635.

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), pp. 120-132. doi: 10.1017/S1355617710001335.

Scherf, K. S., Sweeney, J. A., & Luna, B. (2006). Brain basis of developmental change in visuospatial working memory. *Journal of cognitive neuroscience*, *18*(7), pp. 1045-1058. doi: 10.1162/jocn.2006.18.7.1045.

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology, 41,* pp. 457-471. doi: 10.1007/s10802-012-9684-x.

Seidman, L.J., Biederman, J., Monuteaux, M.C., Valera, E., Doyle, A.E., & Faraone, S.V. (2005). Impact of gender and age on executive functioning: Do girls and boys with and without attention deficit hyperactivity disorder differ neuropsychologically in preteen and teenage years? *Developmental Neuropsychology*, *27*(1), pp. 79-105. doi: 10.1207/s15326942dn2701_4.

Shifrer, D., Muller, C., & Callahan, R. (2016). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. *Journal of Learning Disabilities, 44*(3), pp. 246-257. doi: 10.1177/0022219410374236.

Smith, J.D., Dishion, T.J., Shaw, D.S., Wilson, M.N., Winter, C.C., & Patterson, G.R. (2014). Coercive family process and early-onset conduct problems from age 2 to school entry. *Developmental Psychopathology, 26*(4 0 1), pp. 917-932, doi: 10.1017/S0954579414000169.

Snyder, H.R., Miyake, A., & Hankin, B.L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6,* pp. 1-24. doi: 10.3389/fpsyg.2015.00328.

StataCorp. (2017). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.

Stephens, T. (1981). *Technical manual: Social behavior assessment.* Columbus, Ohio: Cedars Press.

Sulik, M.J., Blair, C., Mills-Koonce, R., Berry, D., & Greenberg, M. (2015). Early parenting and the development of externalizing behavior problems: Longitudinal mediation through children's executive function. *Child Development, 86*(5), pp. 1588-1603. doi: 10.1111/cdev.12386.

Tan, S. (2011). *Internalizing symptoms, executive functions, and academic performance* (Order No. 3444450). Available from ProQuest Central. (858833765). Retrieved from http://ezproxy.alfred.edu:2048/login?url=https://search.proquest.com/docview/858833765?accountid=8263.

Thomas, S.L., & Heck, R.H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education, 42*(5), pp. 517-540. doi: 10.1023/A:1011098109834.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition Guide for Administering and Scoring.* Chicago: Riverside.

Tourangeau, K., Nord, C., Lê, T., Sorongon, A.G., Hagedorn, M.C., Daly, P, & Najarian, M. (2017). *Early childhood longitudinal study, kindergarten class of 2010-2011, user's manual.* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Ursache, A., Noble, K.G., & Blair, C. (2015). Socioeconomic status, subjective social status, and perceived stress: Associations with stress physiology and executive functioning. *Behavioral Medicine, 41,* pp. 145-154. doi: 10.1080/*08964289.2015.1024604.

van Tetering, M. A., & Jolles, J. (2017). Teacher evaluations of executive functioning in schoolchildren aged 9–12 and the influence of age, sex, level of parental education. *Frontiers in Psychology, 8,* pp. 1-10. Doi: 10.3389/fpsyg.2017.00481.

Verhulst, F.C., Koot, H.M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. *Journal of Abnormal Child Psychology, 22*(5), pp. 531-546. Retrieved from http://www.springer.com/psychology/child+%26+school+psychology/journal/10802.

Voyer, D., Voyer, S.D., & Saint-Aubin, J. (2016). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24*(2), pp. 307-334. Doi: 10.3758/s13423-016-1085-7.

Wechsler, D. (1991). *Weschler Intelligence Scale for Children: Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale: Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3^{rd} ed.). San Antonio, TX: Pearson.

Wilkinson, G.S. (1993). *Wide Range Achievement Test—Third Edition*. San Antonio, TX: Psychological Corporation.

Williams, D.R. (1999). Race, socioeconomic status, and health: The added effects of racism and discrimination. *Annals of the New York Academy of Sciences, 896,* pp. 173-188. doi: 10.1111/j.1749-6632.1999.tb08114.x.

Wolfe, R. S., & Renk, K. (2017). Internalizing and externalizing behavior problems in middle school: An ecological-transactional, cross-sectional examination of early adolescent adjustment. *International Journal of Child Health and Human Development, 10*(2), pp. 169–181. Retrieved from http://ezproxy.alfred.edu:2061/login.aspx?direct=true&db=psyh&AN=2018-08781-006&site=ehost-live

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001a). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b). *Woodcock-Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Wootton, J.M., Frick, P.J., Shelton, K.K., & Silverthorn, P. (1997). Ineffective parenting and childhood conduct problems: The moderating role of callous-unemotional traits. *Journal of Consulting and Clinical Psychology, 65*(2), pp. 301-308. doi: 10.1037/0022-006X.65.2.292.b.

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,* pp. 1-9. doi: 10.1016/j.lindif.2012.10.004.

Young, R., & Johnson, D.R. (2015). Handling missing values in longitudinal panel data with multiple imputation. *Journal of Marriage and Family, 77*(1), pp. 277-294. doi: 10.1111/jomf.12144.

Zelazo, P.D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), pp. 297-301. doi: 10.1038/nprot.2006.46.

Zelazo, P.D., Anderson, J.E., Richler, J., Wallner-Allen, K., Beaumont, J.L., & Weintraub, S. (2013). NIH toolbox cognition battery (CB): Measuring executive function and attention. In P.D. Zelazo and P.J. Bauer (Eds.), *National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB): Validation for children between 3 and 15 years. Monographs* *of the Society for research in Child Development*, *78*(4), pp. 16-33. doi: 10.1111/mono.12032.

Zilles, D., Lewandowski, M., Vieker, H., Henseler, I., Diekhof, E., Melcher, T., & ... Gruber, O. (2016). Gender differences in verbal and visuospatial working memory performance and networks. *Neuropsychobiology*, *73*(1), pp. 52-63. doi:10.1159/000443174.

Figure 1. Theoretical Model

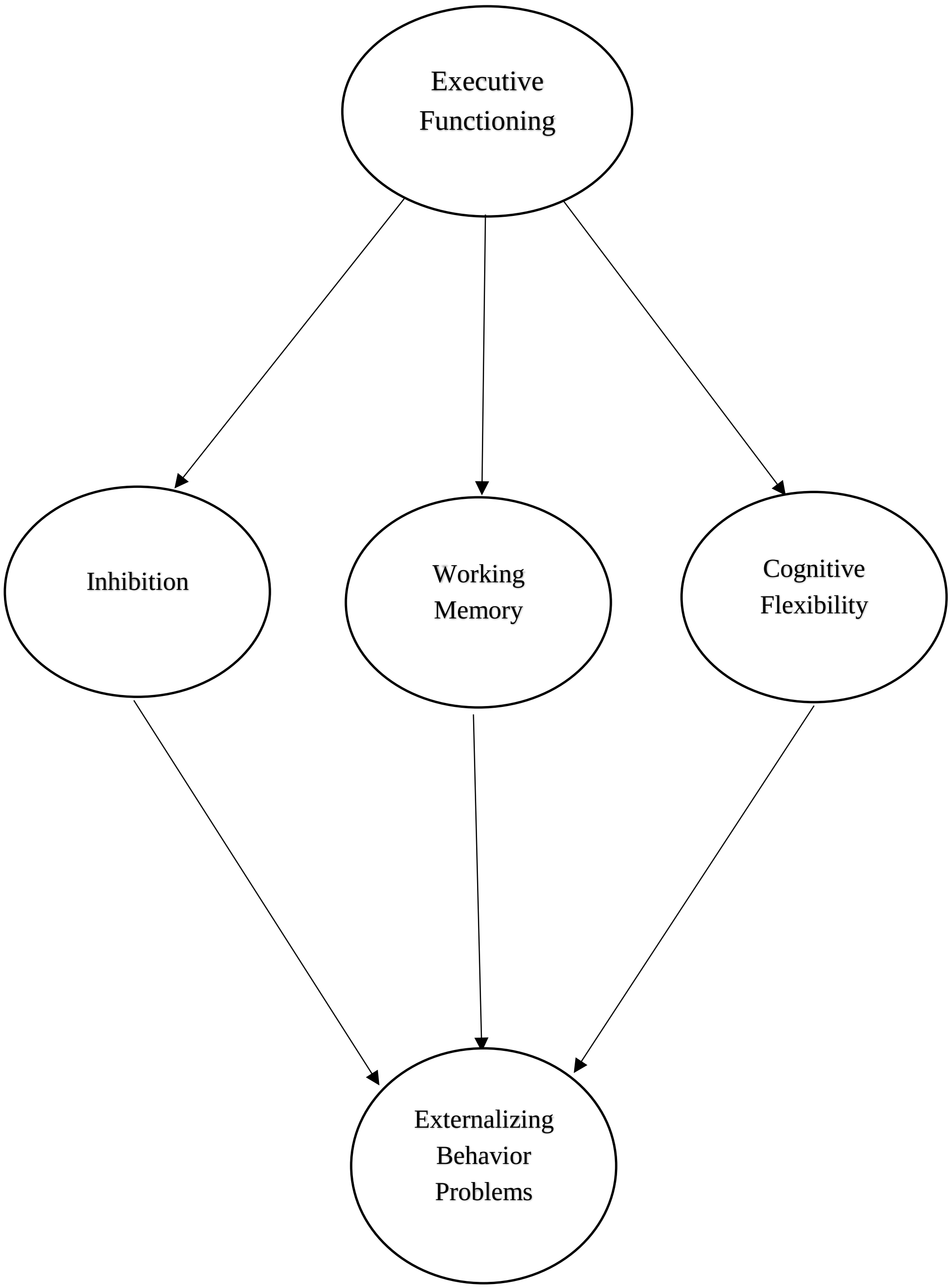

Figure 2. Hypothesized Model

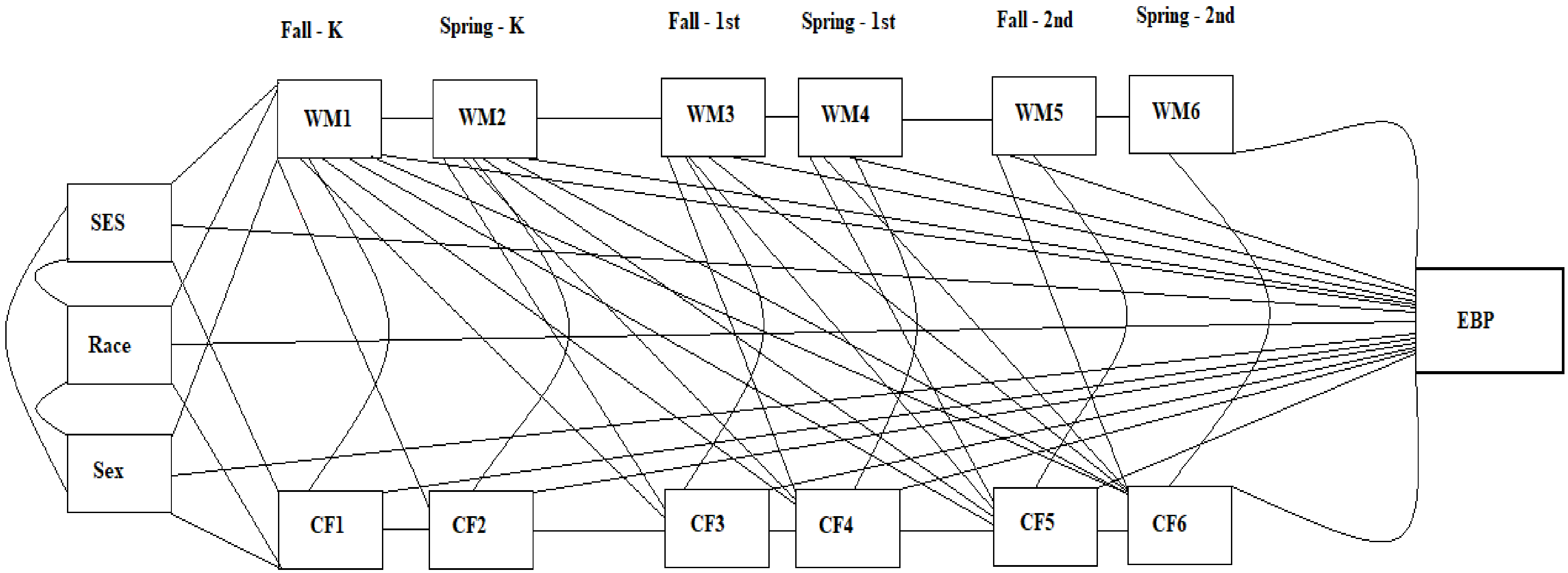

Figure 2. This is the current study as depicted in path format. Single-headed arrows indicate the hypothesized direction of influence while curved lines indicate a correlational relationship. Time points 1 and 2 correspond to the fall and spring of kindergarten year. Time points 3 and 4 correspond to the fall and spring of 1st grade. Time points 5 and 6 correspond to the fall and spring of 2nd grade. Finally, EBP corresponds to teacher ratings of EBP during the spring of 2nd grade. It is important to note that lines from earlier time points of each component of EF to later time points of their respective component are not illustrated in order to keep the figure readable (i.e., WM1 to WM3, WM4, WM5, WM6, etc.). In addition, Race was kept as a single variable, as opposed to the four dummy variables that were used, in efforts to help with readability. Despite this, dummy variables were used during the analysis.

Figure 3. Model Depiction of Main Hypotheses

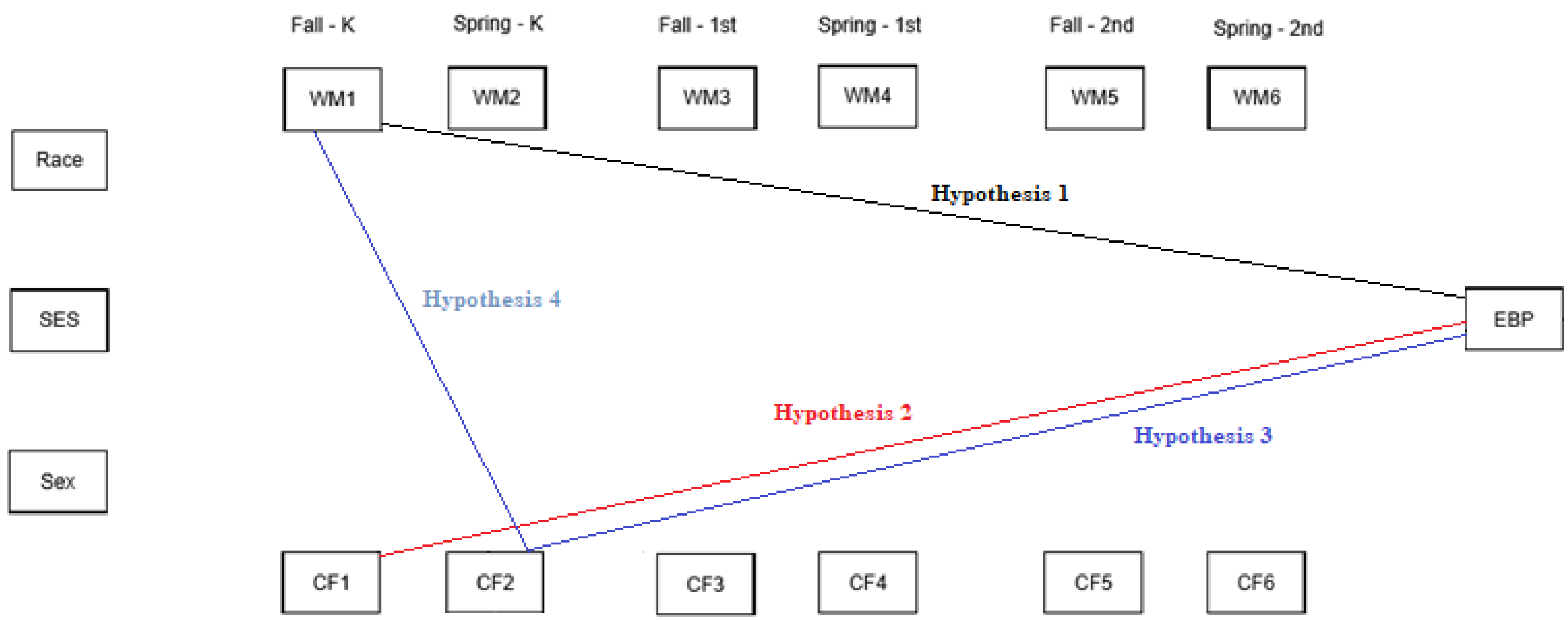

Figure 4. Model Depiction of Standardized Covariances for Control Variables

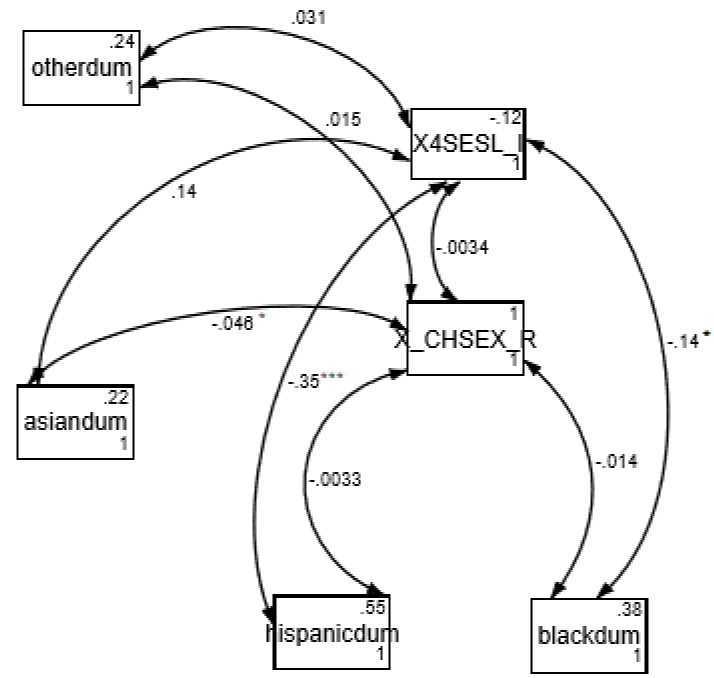

Figure 5. Standardized Covariances among Independent and Dependent Variables

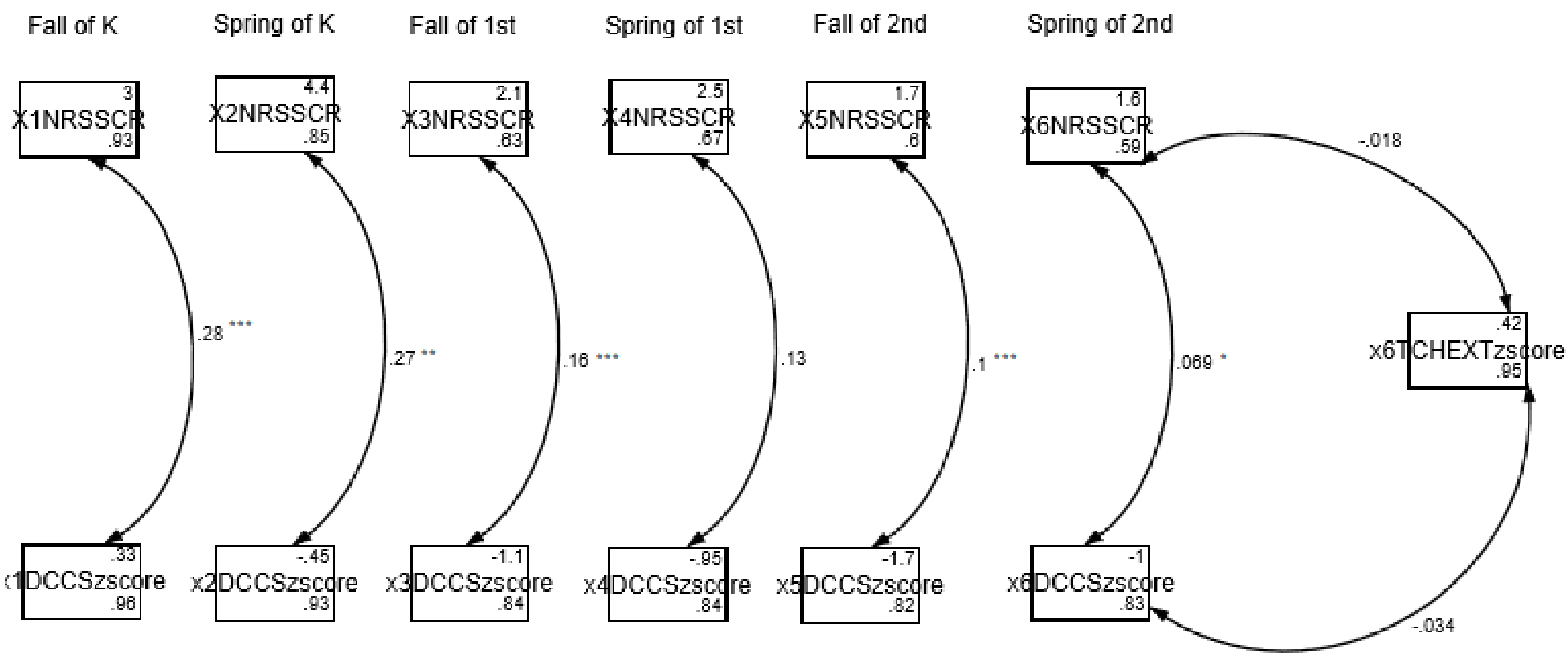

Figure 5. Results of Main Hypotheses

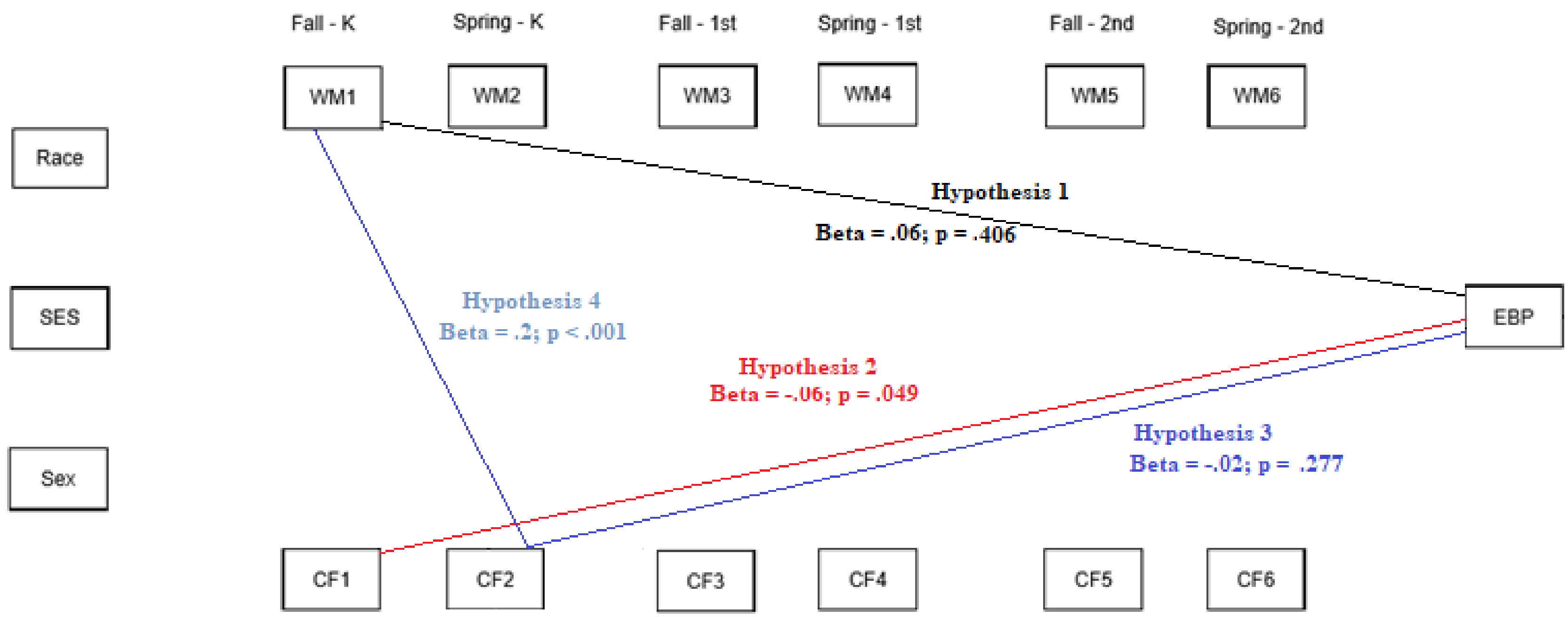

Table 1
Sample Characteristics

Race Classification	Missing	Male	Female	Total
1. White, Non-Hispanic	14	4,403	4,072	8,489
2. Black/African American	8	1,226	1,163	2,397
3. Hispanic, Race Specified	1	2,135	2,048	4,184
4. Hispanic, No Race	12	198	196	406
5. Asian, Non-Hispanic	1	725	817	1,543
6. Native Hawaiian/Pacific Islander	1	59	57	117
7. American Indian/Alaska Native	1	81	86	168
8. Two or More Races	0	433	393	826
Total		9,288	8,847	18,174

Table 2
Descriptive Statistics of Variables

Variable Name	Mean	Standard Error	95% Confidence Interval (CI)	W-Score Mean (SE)	W-Score 95% CI
Control Variables					
SES Composite	-.09	.05	-.20 - .02	-	-
Sex	.51	.01	.49 - .53	-	-
Independent Variables					
WM1*	87.40	1.27	84.83 - 89.98	432.80 (1.72)	429.31 – 436.28
WM2*	95.90	.55	94.79 - 96.99	450.16 (.74)	448.66 – 451.66
WM3*	94.88	.84	93.18 - 96.58	458.73 (1.26)	456.17 – 461.28
WM4*	96.99	.77	95.43 - 98.55	469.99 (1.08)	467.81 – 472.17
WM5*	95.42	.62	94.17 - 96.68	474.05 (.77)	472.49 – 475.61
WM6*	96.72	.65	95.41 - 98.03	480.83 (.83)	479.31 – 482.65
CF1	14.37	.11	14.14 - 14.59	-	-
CF2	15.42	.08	15.26 - 15.58	-	-
CF3	15.92	.08	15.75 - 16.09	-	-
CF4	16.33	.09	16.15 - 16.51	-	-
CF5	6.37	.06	6.25 - 6.49	-	-
CF6	6.94	.04	6.86 - 7.02	-	-
Dependent Variable					
EBP	1.68	.03	1.63 - 1.73	-	-

Note: * indicates Standard Scores were used. Values have been rounded to the nearest hundredth place. In addition, the Race dummy variables are not depicted as

Table 3

Summary of Direct, Indirect, and Total Effects for Externalizing Behavior Problems

Dependent Variable	Independent Variable	Unstandardized Coefficient	Standardized Coefficient (β)	Standardized Indirect Effects	Standardized Total Effects	*t* score	Direct Effect *p* value
EBP	Sex	.28 (.05)***	.14	.00	.14***	5.53	0.000
	SES	-.08 (.02)**	-.06	-.00	-.06*	-3.20	0.003
	Black	.13 (.06)*	.04	.00	.04*	2.24	0.031
	Hispanic	-.14 (.05)*	-.06	.01	-.05*	-2.64	0.012
	Asian	-.29 (.23)	-.06	.00	-.06	-1.28	0.209
	Other Races	.02 (.04)	.01	.00	.01	0.57	0.574
	WM1	.00 (.00)	.06	-.04**	.02	0.84	0.406
	WM2	-.00 (.00)	-.03	-.04	-.07	-0.99	0.326
	WM3	-.00 (.00)	-.04	-.02	-.06	-0.73	0.472
	WM4	.00 (.00)	.02	-.03**	-.01	0.53	0.600
	WM5	-.00 (.00)**	-.07	-	-.07**	-3.52	0.001
	CF1	-.08 (.04)*	-.06	-.00	-.07*	-2.03	0.049
	CF2	-.03 (.02)	-.02	-.00	-.02	-1.10	0.277
	CF3	-.04 (.03)	-.04	.01	-.03	-1.45	0.154
	CF4	.05 (.04)	.04	-.00	.04	1.27	0.213
	CF5	-.00 (.04)	-.00	-	-.00	-0.01	0.988

Note: * = significant at the .05 level; ** = significant at the .01 level; *** = significant at the .001 level.

Table 4

Summary of Direct, Indirect, and Total Effects for Working Memory

Dependent Variable	Independent Variable	Unstandardized Coefficient	Standardized Coefficient (β)	Standardized Indirect Effects	Standardized Total Effects	*t* score	Direct Effect *p* value
WM1	Sex	-2.47 (1.17)	-.04	-	-.04*	-2.10	0.042
	SES	7.31 (.70)***	.18	-	.18***	10.44	0.000
	Black	-11.95 (3.01)***	-.13	-	-.13***	-3.96	0.000
	Hispanic	-7.14 (3.69)	-.10	-	-.10	-1.94	0.060
	Asian	-10.83 (3.07)**	-.07	-	-.07**	-3.53	0.001
	Other Races	-3.89 (3.01)	-.03	-	-.03	-1.29	0.205
WM2	WM1	.22 (.02)***	.38	-	.38***	10.77	0.000
WM3	WM1	.07 (.02)***	.13	.21***	.34***	4.45	0.000
	WM2	.54 (.04)***	.55	-	.55***	14.91	0.000
WM4	WM1	.05 (.01)***	.09	.21***	.30***	4.29	0.000
	WM2	.21 (.02)***	.23	.20***	.43***	8.47	0.000
	WM3	.34 (.04)***	.36	-	.36***	9.29	0.000
WM5	WM1	.00 (.01)	.01	.24***	.25***	0.42	0.679
	WM2	.11 (.02)***	.12	.30***	.42***	4.56	0.000
	WM3	.24 (.02)***	.26	.13***	.39***	9.85	0.000
	WM4	.36 (.03)***	.36	-	.36***	10.81	0.000
WM6	WM1	.00 (.01)	.00	.23***	.23***	0.02	0.983
	WM2	.06 (.03)	.06	.32***	.39***	1.99	0.053
	WM3	.15 (.03)***	.16	.21***	.37***	4.64	0.000
	WM4	.23 (.03)***	.24	.11***	.35***	6.89	0.000
	WM5	.31 (.02)***	.32	-	.32***	13.60	0.000

Note: * = significant at the .05 level; ** = significant at the .01 level; *** = significant at the .001 level.

Table 5

Summary of Direct, Indirect, and Total Effects for Cognitive Flexibility

Dependent Variable	Independent Variable	Unstandardized Coefficient	Standardized Coefficient (β)	Standardized Indirect Effects	Standardized Total Effects	*t* score	Direct Effect *p* value
CF1	Sex	-.11 (.05)*	-.07	-	-.07*	-2.39	0.022
	SES	.10 (.03)***	.09	-	.09***	3.95	0.000
	Black	-.20 (.08)*	-.08	-	-.08*	-2.50	0.017
	Hispanic	-.24 (.06)**	-.13	-	-.13**	-3.78	0.001
	Asian	-.38 (.09)***	-.10	-	-.10***	-4.04	0.000
	Other Races	-.08 (.12)	-.02	-	-.02	-0.69	0.493
CF2	WM1	.01 (.00)***	.20	-	.20***	4.14	0.000
	CF1	.12 (.02)***	.12	-	.12***	4.89	0.000
CF3	WM1	.00 (.00)*	.06	.11***	.17***	2.24	0.031
	WM2	.01 (.00)***	.18	-	.18***	5.49	0.000
	CF1	.13 (.03)***	.12	.03**	.14***	4.35	0.000
	CF2	.24 (.05)***	.22	-	.22***	5.07	0.000
CF4	WM1	.00 (.00)*	.13	.09***	.22***	2.60	0.013
	WM2	.00 (.00)	.05	.08**	.13**	1.32	0.194
	WM3	.00 (.00)**	.08	-	.08**	2.80	0.008
	CF1	.09 (.03)**	.08	.04**	.12***	2.97	0.005
	CF2	.05 (.03)	.05	.05***	.10**	1.61	0.115
	CF3	.22 (.03)***	.23	-	.23***	6.86	0.000
CF5	WM1	.00 (.00)*	.04	.16***	.20***	2.12	0.041
	WM2	.01 (.00)***	.14	.10***	.24***	5.79	0.000
	WM3	.00 (.00)**	.08	.04*	.11**	2.61	0.013
	WM4	.00 (.00)**	.08	-	.08**	3.11	0.003
	CF1	.10 (.03)**	.10	.03***	.13***	3.36	0.002
	CF2	.11 (.03)***	.10	.02**	.13***	4.23	0.000
	CF3	.06 (.03)**	.06	.02***	.09**	2.41	0.021
	CF4	.10 (.02)***	.10	-	.10***	6.84	0.000
CF6	WM1	.00 (.00)	.04	.12***	.16***	1.65	0.107
	WM2	-.00 (.00)	-.00	.16***	.16***	-0.03	0.980
	WM3	.00 (.00)	.03	.09***	.12**	0.91	0.370
	WM4	.00 (.00)***	.08	.05**	.13***	4.24	0.000
	WM5	.00 (.00)**	.09	-	.09**	3.48	0.001
	CF1	.03 (.02)	.03	.05***	.08**	1.49	0.144
	CF2	.02 (.02)	.03	.05***	.07**	1.48	0.147
	CF3	.06 (.02)**	.07	.02*	.09***	3.00	0.005
	CF4	-.00 (.02)	-.00	.03***	.02	-0.22	0.828
	CF5	.25 (.04)***	.26	-	.26***	7.00	0.000

Note: * = significant at the .05 level; ** = significant at the .01 level; *** = significant at the .001 level.