EXPLORING THE STRUCTURE AND CORRELATION OF TEACHER REPORTS AND DIRECT ASSESSMENTS OF LOWER-LEVEL ELEMENTARY SCHOOL CHILDREN'S EXECUTIVE FUNCTIONS

by

BOWEN WANG

(Under the Direction of Kristen L. Bub)

ABSTRACT

The purpose of the current study is to explore whether teacher reports (using the response data of the teacher reports and direct assessments to explore the factor structure of the Behavior Rating Inventory of Executive Function, Second Edition) and direct assessments (using the Early Years Toolbox) adequately measure global executive functions for socio-demographically diverse lower-level elementary school children and to undertake a preliminary investigation of whether a similar structure holds for child sex and grade. Additionally, I sought to understand the relationship between the two measures of global executive functions. Findings suggest that a single global structure of teacher reports holds for the full sample as well as for different samples while a two-factor structure of direct assessments holds for the full sample, for girls, and for each grade. Moreover, teacher reported global executive functions is significantly and moderately correlated with directly assessed global executive functions in the full sample.

INDEX WORDS: Executive functions, Teacher reports, Direct assessments

EXPLORING THE STRUCTURE AND CORRELATION OF TEACHER REPORTS AND DIRECT ASSESSMENTS OF LOWER-LEVEL ELEMENTARY SCHOOL CHILDREN'S EXECUTIVE FUNCTIONS

by

BOWEN WANG

Bachelor of Arts, Southwest University of Political Science and Law, China, 2018

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2020

EXPLORING THE STRUCTURE AND CORRELATION OF TEACHER REPORTS AND DIRECT ASSESSMENTS OF LOWER-LEVEL ELEMENTARY SCHOOL CHILDREN'S EXECUTIVE FUNCTIONS

by

BOWEN WANG

Major Professor:	Kristen Bub
Committee:	Amanda Ferster
	Logan Fiorella

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

ACKNOWLEDGEMENTS

The author would like to thank his committee members, Dr. Amanda Ferster and Dr. Logan Fiorella for their guidance and support throughout this Master's thesis. He would also like to extend his sincerest gratitude to his major professor, Dr. Kristen Bub, for her continual encouragement, dedication and guidance. A special thank you to his family, especially his parents, Shaobo and Lin, his girlfriend, Susan, for their love, understanding, and inspiration.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Research has shown that children with better executive functions are more likely to succeed both academically and behaviorally (Hughes, 1998; Gathercole et al., 2004; Blair & Razza, 2007; Duncan et al., 2007; McClelland et al., 2007; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009; Borella et al., 2010; Morrison et al., 2010; Diamond, 2012). Executive functions are the constellation of higher-order, cognitive processes, and top-down mental processes, which are necessary for concentrating attention, organizing information, and planning goal-directed behavior (Blair & Ursache, 2010; Diamond, 2012; McKinnon & Blair, 2018). Children's executive functions are important for children's learning and development, and they have been measured with both teacher reports and direct assessments. However, there is no clear evidence as to which is a more valid approach to understanding global executive functions. Both teacher reports and direct assessment have their advantages and limitations. The differences in how we measure global executive functions have led to some inconsistencies in the research exploring the effects of these skills on other outcomes like reading and mathematics.

Dimensions and Importance of Executive Functions

Three common dimensions of executive functions were labelled as working memory, inhibition/inhibitory control, and cognitive flexibility (i.e., shifting) (Diamond, 2012; Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003; O'Meagher, Norris, Kemp, & Anderson, 2019). Working memory refers to a mental workplace where information is stored and processed for a short time in the course of demanding cognitive activities, and it enables individuals to temporarily remember information while competitively processing information (Baddeley, 1986; Baddeley & Hitch, 1974; Miyake & Shah, 1999). Working memory can be divided into two domains: visual-spatial working memory and phonological working memory (Baddeley, 1986; Baddeley & Hitch, 1974). Inhibition/inhibitory control involves being able to control one's attention, behavior, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed (Diamond, 2012). Cognitive flexibility is the ability to adapt the cognitive processing strategies to face new and unexpected conditions in the environment (Canas, Quesada, Antoli, & Fajardo, 2003), which involves two aspects: one is being able to change perspectives spatial or interpersonally, the other is being flexible enough to adjust to changed demands or priorities, to admit you were wrong, and to take advantage of sudden, unexpected opportunities (Diamond, 2012). Together, these broad sets of skills are thought to comprise global executive functioning.

Executive functions emerge in infancy and develop throughout the preschool and elementary school age periods sequentially (Carlson, 2005; Diamond, 1990; Garon, Bryson, & Smith, 2008; Hughes, Ensor, Wilson, & Graham, 2010; O'Meagher et al., 2019). Specifically, research indicated that inhibition appears to emerge in infancy and develops rapidly in early childhood, while both cognitive flexibility and working memory develop rapidly between 3 and 8 years of age (Anderson, 2002). It is disproportionately difficult for children in 3-8 age of years to develop inhibitory control, but even infants of 9 to 12 months can start updating the content of their working memory (Diamond, 2012). Children 2.5 years of age can complete the most basic cognitive flexibility task but they cannot fully succeed until 5 years of age (Brooks et al. 2003, Perner & Lang 2002). Therefore, children develop their executive functions mostly in the period of preschool and early elementary school age.

Executive functions play a significant role in children's learning and development. Executive functions in the learning context refer to a kind of skill or attribute that reflects the extent to which children in the classroom persevere with difficult tasks, plan, problem solve, and complete tasks (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009). Children with strong executive functions can control their behavior, mentally manipulate information, and adapt to changing rules and challenges (McClelland et al., 2007; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Also, executive functions are closely associated with children's cognitive characteristics, such as language ability and understanding of false beliefs (Hughes, 1998). Moreover, executive functions contribute to school readiness and school success, especially in numeracy, literacy, and reading competence (Blair & Razza, 2007; Borella et al., 2010, Duncan et al., 2007, Gathercole et al., 2004; Morrison et al., 2010). Children with lower executive functions may have difficulty staying on task, act more impulsively with their peers or teachers, struggle with academic content, or find it difficult to communicate with others in the classroom (Diamond, 2012).

Despite evidence suggesting executive functions are important, the measurement of executive functions remains challenging. There are three general issues. First, many measurements of executive functions focus on one or several separate dimensions of executive functions, especially in direct assessments, but little research focuses on whether the measurements (e.g., teacher reports and direct assessments) are valid to understanding global executive functions. Second, it is difficult to be sure that the three broad sets of skills (e.g., working memory, inhibition, and cognitive flexibility) reflect executive functions, so some reviewers have called for studies attempting to understand executive functions to include multiple measures of these overlapping constructs (Bernier, Matte-Gagné, & Bouvette-Turcot, 2014; Carlson, 2005; Kochanska & Knaack, 2003; Zhou, Chen, & Main, 2012). Third, the construct of executive functions has been defined broadly but measured narrowly, so the distinction between the broad and narrow sense causes the difficulty to measure (Toplak, West, & Stanovich, 2013). In other words, children performing well on a single executive functioning task (narrow) does not suggest that they show competence in global executive functions. To summarize, executive functions are thought to play a significant role on children's learning and development and develop partially as a function of caregiver socialization, so measurement is important to consider (Bari & Robbins, 2013; Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Conway & Stifter, 2012; Hammond, Müller, Carpendale, Bibok, & Liebermann-Finestone, 2012). In this paper, two measures of executive functions involving one teacher report and one direct assessment with their advantages and limitations are discussed in the following sections.

Teacher Report

Teacher reports can be considered as an important source for measuring children's executive functions. Teacher reports of children's executive functions include measures like the Work-Related Skills (WRS) and Interpersonal Skills (IPS) subscales of the Cooper-Farran Behavioral Rating Scale (CFBRS; Cooper & Farran, 1988, 1991), the attention focusing and inhibitory control subscales of the Children's Behavior Questionnaire (Rothbart, Ahadi, Hershey, & Fisher, 2001), Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008; Thorell et al., 2010), Working Memory Rating Scale (WMRS; Alloway et al., 2008), Dysexecutive Questionnaire (DEX); Part of the Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996) and Behavior Rating Inventory of Executive Function (Gioia, Isquith, Retzlaff, & Espy, 2000). Moreover, a teacher survey was designed especially for

K-5th grade children's self-control, persistence, and social competence (Scarupa, 2014). Additional information of each of these teacher reports of executive functions is listed in Table 1.

Advantages. Researchers believe that teacher reports have advantages for measuring children's executive functions. Specifically, teacher reports are relatively cost efficient compared to direct assessments because they typically occur in the classroom (Garcia, Sulik, & Obradović, 2019). Moreover, teacher reports of executive functions can provide unique information about students' self-regulation capacities in the context of the classroom environment that is not reflected in direct assessments (Fuhs, Farran, & Nesbitt, 2015) or even parents reports. Toplak et al. (2013) argued that teacher reports are believed to have ecological validity because they capture children's behavior in real-world contexts, which suggests that teacher reports capture how children use executive function skills in the dynamic and distracting classroom environment. Furthermore, researchers indicated that teachers' reports of executive functions can independently predict academic achievement (Blair, Ursache, Greenberg, & Vernon-Feagans, 2015; Fuhs et al., 2015; Obra-dović, Sulik, Finch, & Tirado-Strayer, 2018). Hence, the advantages of teacher reports include cost efficiency, ecological validity, and more utility in measuring students' academic achievement in the classroom environment.

Limitations. Although teacher reports of executive functions can have powerful implications for children's academic skills and cognitive ability, they can also be influenced by some external factors including teachers' socioeconomic status and students' demographic characteristics; as such teachers' perceptions of executive function skills may be biased (Garcia, Sulik, & Obradović, 2019; Nisbett & Wilson, 1977; Ready & Wright, 2011). The bias refers to the degree to which teachers systematically overestimate or underestimate students' skills, controlling for a direct assessment. For example, Ready and Wright (2011) argued that teachers in lower-socioeconomic-status and lower-achieving contexts may more often underestimate their students' cognitive abilities. Also, student-teacher relationship quality biases teachers against students' skills (Hughes, Gleason, & Zhang, 2005). Furthermore, because there is evidence that academic skills are strongly related to executive functions (Matthews, Ponitz, & Morrison, 2009; Ursache, Blair, & Raver, 2012), by evaluating whether children's academic skills (e.g., literacy skills and reading skills) are affected by the bias of teacher reports, Garcia et al. (2019) estimated whether children's executive functions measured by teacher reports are also biased. Research indicated that teacher reports on children's academic skills were inconsistent with direct assessments of children's academic skills because teachers were influenced by children's various demographic characteristics including gender, ethnicity, and ELL status (Baker, Tichovolsky, Kupersmidt, Voegler-Lee, & Arnold, 2015; Hinnant, O'Brien, & Ghazarian, 2009; Hughes, Gleason, & Zhang, 2005; McKown & Weinstein, 2008; Ready & Wright, 2011). Therefore, similar to the relationship between teacher reports of children's academic skills and teachers' bias from children's demographic characteristics, Garcia et al. (2019) concluded that teacher reports of children's executive functions may also be influenced by the teachers' bias based on children's gender, ethnicity, and ELL status.

The other limitation of applying teacher reports is the structure of teacher reports, which may affect the validity of the measures. To be more specific, researchers questioned the underlying constructs of one of the most famous teacher reports–BRIEF (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007, Fuhs and Day, 2011, Spiegel, Lonigan, & Phillips, 2017). For example, Duku and Vaillancourt (2014) questioned the validity of the preschool version of BRIEF (BRIEF-P) structure, and they reported that some subscales of it could be separated into further categories, such as the Inhibition subscale into "awareness" and "impulsivity," and the

Shifting subscale into “inflexibility,” “adjusting,” and “sensory”. Moreover, Spiegel et al. (2017) conducted factor analyses with 2,367 preschool students and indicated that the subscales titles (e.g., Inhibition, Shifting, Emotional Control, Working Memory and Planning/Organizational skills) in BRIEF-P did not reflect the subscales items enough. Thus, the underlying constructs of teacher reports as one of the limitations may affect the validity of the measures for different population.

Direct Assessment

Direct assessments refer to methods of measuring executive functions using standardized tasks that place demands on children’s executive function skills. The measures are typically based on the participants’ accuracy, response time, and/or speed responding under a limit time (Toplak, West, & Stanovich, 2013). Often, direct assessments are performance-based. For example, participants need to focus on target stimuli when ignoring distractors or remember strings of digits and then report them backward (Garcia, Sulik, & Obradović, 2019). There are some direct assessments for measuring executive functions, such as the Wisconsin-Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtis, 1993) and the Stroop test (Jensen & Rohwer, 1966; MacLeod, 1991; Stroop, 1935). Particularly, the WCST requires the maintenance of a task set, flexibility in response to feedback, avoiding perseverative tendencies, and inhibiting a prior response that is no longer appropriate (Salthouse et al., 2003). The Stroop test is to measure children’s interference control (cognitive flexibility). Children must inhibit an overlearned response (reading a word that names a color) to respond with another dimension that is incongruent and ‘interfering’ (naming the ink color of the word, instead of the actual color word). The National Institute of Health (NIH) Toolbox (NIH Toolbox CB, 2013) is a freely available and widely accessed measure for assessing cognitive (e.g., EF, attention, memory), emotional (e.g., well-being, stress), motor (e.g., locomotion, strength), and sensory abilities (e.g., audition, vision) key aspects of development from 3 to 15 years of age (Zelazo et al., 2013). There are a number of direct assessments of executive functions and several examples are presented in Table 2.

Advantages. Direct assessments also have advantages for measuring children's executive functions. Silver (2014) believed that direct assessments are the most objective measure of executive functions. In other words, direct assessments are not affected by bias such as teachers' perceptions or some external interruption(s). The fact is important because teachers' bias is detrimental to students' academic performance and self-efficacy. For example, teachers' bias can be based on gender. Gender bias can lead teachers to believe girls' academic performance is inferior to boys' in STEM fields (Fennema et al., 1990; Hand, Rise, & Greenlee, 2017), although there are negligible differences between men and women in innate intellectual aptitude (Halpern, 2013). Indeed, girls may perform better academically than boys in mathematics and science courses (Shettle et al., 2007). This bias may prevent girls from further exploring their interest in STEM fields. Given that there is evidence that academic skills are strongly related to executive functions (Matthews et al, 2009; Ursache et al, 2012), it is necessary to avoid biased measures of executive functions. Therefore, in terms of bias reduction, direct assessments are better measures of executive functions. Furthermore, direct assessments can capture children's basic cognitive processes and assess information processing mechanisms such as working memory and inhibition, something that can be difficult to capture with teacher reports (Toplak, West, & Stanovich, 2013).

Limitations. One of the challenges in the assessment of executive functions is the impurity problem (Toplak, West, & Stanovich, 2013). That is, executive functioning tasks can not only tap non-EF skills but also typically focus on more than one executive functioning skill (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). In other words, most direct measures of executive functions involve non-executive processes in the task context, such as color naming in the Stroop task (Miyake & Friedman, 2012). Howard and Melhuish (2017) argued some limitations of Early Year Toolbox (EYT). Given that there are only four tasks associated with executive functions in the EYT, researchers cannot fully evaluate the dimensionality of executive functioning using the EYT response data. Because the “task impurity” issue exists commonly in measuring executive functions, the latent variable data analysis method (e.g., confirmatory factor analysis and structural equation modeling) is commonly conducted to minimize the likelihood of conclusions that are influenced by variance unrelated to the constructs of interest (Friedman & Miyake, 2004; Howard Johnson, & Pascual-Leone, 2014; Miyake et al., 2000). Since there are only four sub-measures in the EYT (inhibition, shifting, and two measures of working memory), a latent variable approach will be more valid if complementary measures from outside the EYT (e.g., teacher reports and parent reports) can be used. Regardless, it is still necessary to develop more age- and sample- appropriate tasks (Howard & Melhuish, 2017). Besides that, it is sometimes impractical to conduct direct assessments. For example, when implemented in school settings, some direct assessments typically take place one-on-one in a quiet room separate from the classroom. Also, direct assessments take longer time than teacher reports, especially in individually administered direct assessments requiring more people and time resources. Even in the more advanced EYT tasks, every task takes a child about 5-8 minutes (Howard & Melhuish, 2017). Four executive functioning tasks take a child about half an hour totally, while teacher reports (e.g., BRIEF) takes 10-15 minutes to measure a child (Gioia et al, 2000). Hence, this approach is sometimes expensive, time-intensive, and impractical for large-scale data collection (Obradović et al., 2018).

Correlation between Teacher Reports and Direct Assessments

Although teacher reports and direct assessment are both commonly used, many researchers believe the association between teacher reports and direct assessments is low. Indeed, for various populations the correlations typically range from .10 to .30. The populations of these studies contain both children and adults (Toplak, et al., 2013), 6-15 years of age children (Mcauley, Chen, Goos, Schachar, & Crosbie, 2010), preschool children (Clark, Pritchard, & Woodward., 2010; O'Meagher, Norris, Kemp, & Anderson, 2019; Sulik et al., 2010), and both preschool and early elementary school children (Blair et al, 2015). For example, Toplak et al. (2013) scrutinized 20 studies and argued that the performance-based assessments (i.e., direct assessments) and rating measures (i.e., teacher reports) of executive function generally have a low association (overall median correlation was only .19).

There are some explanations for this inconsistency. The first primary reason suggested by Toplak et al. (2013) is that the two measures of executive functions assess different underlying mental constructs. Direct assessments of executive functions provide important information about the efficiency of cognitive abilities (i.e., algorithmic level of analysis) (Anderson, 1990; Marr, 1982; Stanovich, 1999, 2009). However, teacher reports of executive functions appear to capture the success in rational goal pursuit. In other words, the rational goal pursuit can be explained as people's choice of action that is rational given the beliefs relevant to people's goals (Bratman, Israel, & Pollack, 1988; Dennett, 1987; Newell, 1982, 1994; Pollock, 1995; Stanovich, 2009, 2011). Teacher reports focus more on the choice of action to achieve goals and decision-making while direct assessments focus on information-processing mechanisms in the brain (e.g., working memory and inhibition). Moreover, as described above, executive functions are essentially difficult to measure accurately because the underlying constructs of executive functions have commonly been defined broadly but measured narrowly. It is not surprising then that correlations among direct assessments and teacher reports are low.

The other reasons of the low association between teacher reports and direct assessments are about their advantages and limitations, which generally include a) direct assessments involve considerable structure and direction from the examiner, whereas teacher reports as one of the rating measures involve very little direction from the examiner (Toplak et al., 2013); b) similar to assessing children's academic skills, teacher reports on children's executive functions are inconsistent with direct assessments because children vary systematically according to their gender, ethnicity, and ELL status (Garcia et al., 2019); and c) teacher reports of executive functions can be independent on direct assessment to uniquely predict children's academic achievement (Blair et al., 2015; Fuhs et al., 2015; Obra-dovic ́et al., 2018).

The ideal condition is that teacher reports can be considered accurate and unbiased when their results correspond to direct assessments of student skills. Even though much research indicated the low association between teacher reports and direct assessments, a small number of studies showed a stronger association between teacher reports and direct assessments. For example, Fuhs et al. (2015) used some teacher reports (e.g., Cooper-Farran Behavioral Rating Scale, CFBRS as the teacher report) and direct assessments (e.g., Corsi Blocks task, Dimensional Change Card Sort, DCCS, Copy Design, Peg Tapping, and Head Toes Knees Shoulders, HTKS) to study 719 prekindergarten children. They concluded that the direct assessments and teacher reports of executive functions (i.e., Work-Related Skills, WRS) were significantly and moderately correlated (the range of *r* is from .257 to .511), which implies that they measure, in part, a similar underlying construct.

Current Study

As the literature suggests, in most cases, teacher reports and direct assessments of executive functions are inconsistent and not strongly correlated. The reason for the inconsistency is related to the characteristics of the two measures including their advantages and limitations. Both teacher reports and direct assessments have their limitations. The limitations of teacher reports of executive functions mainly contain the bias from teachers' perceptions and the uncertainty of underlying constructs while the limitations of direct assessments include task impurity and practicality concerns. And the task impurity is caused by the uncertainty of underlying constructs of measures. Considering these limitations and the issue that the construct of executive functions was defined broadly but measured narrowly, the ecological validity of executive functioning measures comes into question. In addition, most research focused on the measurements for broad sets of executive function skills (working memory, cognitive flexibility, or inhibition) and the correlation among teacher reports and direct assessment in each domain of executive functions. In other words, they did not concentrate on global executive functions.

Thus, the purposes of the current study include a) explore whether teacher reports and direct assessments adequately measure global executive functions for socio-demographically diverse lower-level elementary school children; b) a preliminary investigation of whether a similar structure holds by sex and grade; and c) understand the relationship between the two measures of global executive functions for socio-demographically diverse lower-level elementary school children. To be more specific, I used the Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2; Gioia et al, 2015) as the measure of teacher reported executive functions, which is an assessment of executive function behaviors at school for children and adolescents. In addition, the direct assessments of children's executive functions used in this study are from the Early Years Toolbox (EYT).

The research questions are listed as follow:

A) Through the preliminary analysis, is it reasonable to assume that the expected constructs of teacher reports (three subscales versus one global scale) and direct assessments (two subscales versus one global scale) hold in a diverse sample of lower-level elementary school children?

B) Within this sample, which tool can more adequately measure global executive functions between teacher reports or direct assessments?

C) Within this sample, to what degree are the teacher reports and direct assessments of global executive functions correlated?

CHAPTER 2

METHODS

Sample

Data from the *National Institutes of Health (NIH)* study of *Links between Classroom Daylighting, Sleep and Learning-Related Skills (LCDSLS)* were used for the current study. One hundred and sixty-two students from three cohorts of diverse 6 to 10 years old from 13 classrooms (1^{st}, 2^{nd}, or 3^{rd} grade) were recruited. To maximize variability in the amount of daylight exposure, the *LCDSLS* researchers recruited schools from urban and rural communities in the Midwest with varying numbers and styles of classroom windows. All 1^{st}, 2^{nd}, and 3^{rd} grade teachers within schools that consented to participate were invited to take part in the study. The researchers also invited all children and their families of participating teachers to be in the study. Children were invited to participate for one academic year while teachers were asked to participate for two. This cohort design allowed the researchers to replicate analyses with distinct samples that have been exposed to similar classroom experiences. I performed a secondary data analysis based on the *LCDSLS* research.

The current study is based on a sample of 162 lower-level elementary school children attending a rural or urban school in the mid-west. However, 12 students are either absent from the direct assessments or missing some items in the teacher reports. Therefore, a sample of 140 students are analyzed in this research. Of the sample, participants' age ranged from 6 to 10 years with mean age equal to 7.17 years; 45% of the children were female (N = 63, coded as 1), 55% of the children were male (N = 77, coded as 0), 42.1% were 1^{st} grade (N = 59), 42.9% were 2^{nd}

grade (N = 60) , 15% were 3rd grade (N = 21). Also, of the sample, 61.4% of the children were white (N = 86), 5% were black of African American (N = 7), 5.7% were two or more races (N = 8), 7.9% were other race/ethnicity (N = 11), 20% data of children's race/ethnicity were missing. And we knew there were 16 Latino children. The age of children's moms ranged from 26 to 57 years and mean age was 35.62 years. Except for 26 missing data, about 42% mothers had bachelor's or higher degree (N = 47), only 2.6% were educated less than high school (N = 3). Except for 31 missing data about family annually income, the children's family annually income was various from less than $5,000 to more than $300,000. About 17% of family (mode group) had annually income from $100,001 to $150,000 ($N$ = 24). And the average annually income was from $60,001 to $65,000. Thirteen teachers were recruited to participate the study to complete the teacher reports.

Procedure

Participant enrollment and data collection were conducted are part of the *LCDSLS* research. Teachers were invited to participate in the projects. Those who agree sent home information letters and consent forms to all children in their classes. Two weeks after the initial letter went home, teacher send a reminder about the study and one week later, data collection began. Demographic data were collected via parent questionnaires sent home by the teachers. Theses questionnaires were return to teachers in a sealed envelope to maintain confidentiality. Teachers completed complete the BRIEF2 questionnaire on each participating child in their classroom by the end of the fall or spring semester (depending on when they were participating). Trained observers collected the direct assessment data in late fall or late spring (again depending on the cohort). More specifically, students were pulled from regular class activities in groups of 3-5 and were given their own iPad to complete the Early Tears Toolbox activities. Following the completion of each game, children indicated to the research assistant they were ready to move on and were then started on another game. Games were administered in random order to ensure that participant fatigue (if there was any) did not affect any single game. All protocols were approved by the participating institutions' Human Subject's boards.

Measures

BRIEF2. The BRIEF2 (Gioia et al, 2015) published by PARInc., is an individually administered rating scale of executive function for children and youth, aged 5 to 18 years. This study used the teacher form as the measure of teacher-reported executive function problems. Teachers rated each of 63 items using a Likert-type scale with N ("Never", adding 0 point), S ("Sometimes", adding 1 point), or O ("Often", adding 2 points). Sample items include: "when given three things to do, remembers only the first or last", or "become upset with new situation". The items are used to create theoretically and statistically derived scales measuring different aspects of children's behaviors, such as their ability to control impulses, move freely from one situation to the next, modulate responses, anticipate future events, and keep track of the effect of his or her behavior on others. More specifically, the BRIEF2 items can be combined to represent three components of executive functioning including the Behavior Regulation Index (BRI), Emotional Regulation Index (ERI), and Cognitive Regulation Index (CRI). Together, these three subscales form the Global Executive Composite (GEC). The higher a child is rated on each index or on the global executive function scale, the more problems teachers perceive the child to have in that domain.

The BRIEF2 was developed based on the standardization sample consisting of 3,603 children and the clinical sample consisting of 5,295 participants. There were 1,400 children in the standardization sample and 1, 826 children in the clinical sample contributing to the teacher form (Gioia et al, 2015). The standard sample were typically developing, which meant that the sample was required to have no history of special education, psychotropic medication, or neurological disorders. In addition, teachers were required to know the student through daily contact for at least 1 month. Clinical samples included Attention-Deficit/Hyperactivity Disorder Combined Presentation (ADHD-C), ADHD Predominantly Inattentive Presentation (ADHD-I), sluggish cognitive tempo, Autism Spectrum Disorder, learning disability (LD), comorbid ADHD and LD, anxiety, traumatic brain injury, epilepsy, neurofibromatosis, cancer, or diabetes (Gioia et al, 2015). Based on the author's data, findings indicate high internal consistency for all index scores in both the standardization and clinical samples, which revealed coefficients ranging from .88 to .98, with index and composite scores ranging from .94 to .98. Also, the test-retest reliability is also tested by correlational analysis between two administrations approximately 3 weeks apart, with the range from .67 to .92 (M = .79). Findings also reflect the sources of validity including content and internal structure. The internal structure was measured by examining item-total correlations, intercorrelations, and confirmatory factor analysis. The results of these analyses indicated the good internal structure of the BRIEF2. For example, the item-total correlation coefficient ranged from .50 to .83 and the intercorrelation coefficient ranged from .46 to .88. Moreover, studies have ascertained the goodness-of-fit of a three-factor structure (Behavior Regulation, Emotion Regulation, and Metacognition) for the original BRIEF via exploratory and confirmatory factor analyses, and the CFI was .95, with an SRMR of .04 (Gioia, Isquith, Retzlaff, & Epsy, 2002).

Early Years Toolbox (EYT) tasks. Four tasks in EYT (or five tasks if consider the "Go/No-Go" task as two subtasks) were used to directly assess children's executive functioning skills. The EYT is a collection of iPad-based short game-like tasks designed to measure young children's emerging cognitive, self-regulatory, language and social development (Howard & Melhuish, 2017). The four assessments used in the current study are "Mr. Ant" (measuring visual-spatial working memory), "Not This" (measuring phonological working memory), "Go/No-Go" (measuring inhibition), and "Card Sorting" (measuring shifting). In other words, the "Go/No-Go" tasks measure behavior regulation and the other three tasks measure cognitive regulation. The four measures are developmentally appropriate and sensitive, brief, engaging, valid, and reliable for use with young children. all tasks are designed to terminate after a set number of incorrect responses (depending on the task) so as not to cause frustration in the children. For all dimensions, higher scores indicate better executive function skills. The EYT demonstrates high internal consistency and is moderately correlated with the Flanker task, Dimensional Card Sorting Task, NIH Toolbox, and Strengths and Difficulties Questionnaire (Howard & Melhuish, 2017).

"Mr. Ant" Task. The "Mr. Ant" Task is adapted from Case's (1985) Mr. Cucumber task and following the protocols of Morra (1994). The task is an iPad-based assessment of "visual-spatial working memory" which indicates the amount of visual information that concurrently can be coordinated in mind. In this task, children are presented with an image of a cartoon character—Mr. Ant—who has a number of colored stickers placed in different parts of his body. After a predetermined amount of time, these dots disappear, and the child is then asked to recall the locations of the dots by tapping the spatial locations on Mr. Ant that they believe previously held stickers. Test trials increase in difficulty (i.e., working memory demand) as the task progresses, with three trials at each level of complexity. The task consists of eight levels (24 trails total), progressing from one to eight stickers. Each trail is organized in the following steps: 1) Mr. Ant is presented with *n* colored stickers (where *n* equals the current level of difficulty) for

5 seconds; 2) a blank screen is presented for 4 seconds is presented; and 3) an image of Mr. Ant without stickers—along with an auditory prompt to recall where the stickers were—is presented until the participant's response is complete. The task continues until either the completion of all 8 levels or failure on three trials at the same level of difficulty, whichever occurs first. The scoring rule is to calculate the number of correct trails beginning from level 1. This task takes approximately 5-10 minutes to complete, depending on the student's level of development.

"Not This" Task. The "Not This" Task is loosely based on the Direction Following Task (Im-Bolter, Johnson, & Pascual-Leone, 2006). This task is an iPad-based assessment of "phonological working memory" (i.e., the amount of auditory information that concurrently can be coordinated in mind). In this task, children are presented with an array of characters that vary in shape, size and color (e.g., a large blue circle). Before the characters are shown, participants hear an auditory instruction to point to a card that does not fulfill a certain descriptive quality. The task consists of five trials at each level of complexity (8 levels and 40 trails total), the difficulty of which is aligned with the number of stimulus features that must be concurrently activated in mind. For each level, the number of descriptive qualities that must be held in mind increases. For example, an item at Level 1 trial may be "Find a shape that is not blue" (one feature—blue), whereas an item at Level 3 trial may be "Find a shape that is not small, not red and not a triangle" (three features—small, red, triangle). Each trail is organized in the following steps:1) an auditory instruction played against a white screen; 2) a 3-second interval between the introduction and characters presented on the screen; and 3) a 4 × 5 array of various colored and sized shapes with cartoon faces, presented until a response is made by tapping the shape(s) that the participant believes correspond to the auditory instruction (Howard & Melhuish, 2017). The task continues until either the completion of all 8 levels or failure on three trials at the same level of difficulty, whichever occurs first. The scoring rule is to calculate the number of correct trails beginning from level 1. This task takes approximately 5-10 minutes to complete, depending on the student's level of development.

"Go/No-Go" Task. The Go/No-Go Task is based on previous protocols such as Fish-Shark go/no-go task (Howard & Okely, 2015; Wiebe, Sheffield, & Espy, 2012). This task is an iPad-based assessment of inhibition (the ability to control behavioral urges). In this assessment, children are presented with fish and sharks and are instructed to tap the iPad screen whenever they see a Fish (catch the fish) and refrain from responding when a Shark appears (avoid the sharks). Since there are 70% go trails as the majority of stimuli, which generates a prepotent tendency to response (impulse), it requires children to inhibit the response on 30% no-go trails. The task proceeds with 50-75 stimuli divided evenly into three test blocks (each separated by a short break and a reiteration of instructions). Stimuli are presented in pseudo-random order. That is, a block never begins with a no-go stimulus and no more than two successive trials are no-go stimuli (Howard & Melhuish, 2017). The scoring procedure is to calculate the product of proportional "go" (to account for the strength of the prepotent response generated) and "no-go" accuracy (to index a child's ability to overcome this prepotent response). This game takes approximately 5 minutes to complete. Howard & Melhuish (2017) conducted internal consistency analyses and concluded that "Go/No-Go" had good reliability for both go (Cronbach's $\alpha = .95$) and no-go trials (Cronbach's $\alpha = .84$) among 1,764 preschool and early primary school students.

"Card Sorting" Task. The Card Sorting Task is based on the protocols of Zelazo (2006). This task is an iPad-based assessment of "shifting" (an executive function that involves the ability to control and redirect attention). In this assessment, after a demonstration trial and two practice trials, children are presented with cards that vary along two dimensions (i.e., shape and color) and are asked to sort each card (i.e., red rabbits and blue boats) first by one dimension (e.g., color) as the pre-switch stimuli and then, after six trials, by another dimension (e.g., shape) as the post-switch stimuli. If children correctly sort at least five of the six pre- and post-switch stimuli, they proceed to a border phase of the task. In this phase, children are required to sort by color if the card has a black border or sort by shape if the card has no black border (Howard & Melhuish, 2017). In all conditions, each trial begins by reiterating the relevant sorting rule and then presenting a stimulus for sorting. The scoring rule is to calculate the number of correct trails beginning from the first trail. Their ability to flexibly shift from one sorting rule to another corresponds to their cognitive flexibility (or shifting). This task takes approximately 5-7 minutes to complete, depending on the student's level of development.

Analysis Plan

Descriptive analyses. I conducted descriptive statistics to obtain the means and standard deviations of the scores in both the teacher report (i.e., BRIEF2) and the direct assessment (i.e., four EYT tasks) for the total analytic sample. An analysis of the distributional properties for each composite was also conducted. Furthermore, to better understand the characteristics of the sample, I considered the descriptive statistics for each sex and each grade.

Principal Components Analyses (PCA). I conducted several PCAs to answer the research questions. The goal of a PCA is to replicate the correlation matrix using a set of components that are fewer in number and linear combinations of the original set of items (UCLA: Statistical Consulting Group). PCA makes no assumption about a model and is only concerned with which linear relationships exist and how any particular variable might contribute to that relationship. In other words, the purpose of PCA is to assist in scale development—as opposed to structure confirmation. It is typically used for data reduction and understanding whether items relate to intended components. In the current study, the data reduction was to identify a smaller set of components than number of items for parsimony sake. All of the PCAs were conducted via IBM SPSS 24 version. For the purpose of exploring whether the suggested structures of the teacher reports and direct assessments hold in a diverse sample of lower-level elementary school children, I conducted two separate PCAs: one for the teacher report (i.e., BRIEF2) and one for the direct assessment (i.e., EYT tasks) each using the full sample. Specifically, for the BRIEF2, I conducted PCA to explore whether the three broadband dimensions of executive functioning (i.e., BRI, ERI, and CRI) hold as a three-factor structure in this sample of lower-level elementary school children. The three-factor structure would hold in the sample, if there are three components extracted, which is as same as the number of subscales the authors suggested (three subscales including BRI, ERI, and CRI). For the EYT tasks, I conducted a PCA to explore whether two broad domains (i.e., cognitive regulation and behavior regulation) hold as a two-factor structure for the sample of lower-level elementary school children. The two-factor structure would hold in the sample if there are two components extracted, which is as same as the number of dimensions based on the theory of executive function (two dimensions including cognitive and behavior regulation). Then, I essentially did the same thing to explore whether a single global executive function score holds for each measure for the full sample. Finally, as a first step in considering whether the teacher reports or direct assessments of global executive functions look similar across different subsamples (i.e., males and females or 1st, 2nd, and 3rd graders), I fit separate PCAs for each subsample.

Before conducting PCA, Kaiser-Meyer-Olkin (KMO) and Bartlett’s test is necessary to determine if the data are appropriate for PCA. The Kaiser-Meyer-Olkin Measure of Sampling

Adequacy is a statistic that indicates the proportion of variance in your variables that might be caused by underlying factors. Kaiser and Rice (1974) suggested that the KMO must be greater than .50 even though it is still miserable in the .50s. It is mediocre in the .60s, middling in the .70s, meritorious in the .80s, and marvelous in the .90s (Kaiser & Rice, 1974). Bartlett's test of sphericity tests the hypothesis that the correlation matrix is an identity matrix in large sample size, which would indicate that the variables are unrelated and therefore unsuitable for structure detection. Small values (less than 0.05) of the significance level indicate that a factor analysis may be useful with the data. The sample size of this study was more than one hundred, which was sufficient for Bartlett's test. These two criteria are to determine whether PCA can be processed.

Then, to determine how many components should be selected, the criteria of eigenvalue is necessary. Eigenvalues represent the total amount of variance that can be explained by a given principal component. Also, eigenvalues are the sum of squared component loadings across all items for each component, which represent the amount of variance in each item that can be explained by the principal component. The eigenvalue should be greater than 1 in order to retain the factor (Kaiser, 1960). Finally, factor loading represents whether the factor (component) extracts sufficient variance from that variable. Tabachnick, Fidell, and Ullman (2007) cite the minimum loading of an item is .32, which equates to approximately 10% overlapping variance with the other items in that factor. And .50 or higher factor loading indicates strong loaders on the factors (Costello & Osborne, 2005).

Correlational analyses. Once the preliminary factor structure was determined, I created two separate weighted averages reflecting global executive functioning for either the teacher report or global executive functioning for the direct assessment. I did this by using the factor loadings of the BRI, ERI, and CRI subscales in the BRIEF2 and the five subscales (i.e., “Mr. Ant”, “Not This”, “Go”, “No-Go”, and “Card sorting”) in the EYT. Specifically, I multiplied the factor loading and the raw score of each subscale to obtain the weighted scores.

Then, I added the three weighted subscales’ scores in the BRIEF2 to obtain the weighted global score of the teacher report. Likewise, I added the five weighted subscales’ scores in the EYT to obtain the weighted global score of the direct assessments. Using these global weighted scores, I conducted a bivariate correlational analysis to explore whether teacher reports and direct assessments of children’s global executive functions are correlated in this sample. Additionally, to further explore these correlations by gender and grade, I again conducted five separated bivariate correlational analyses (for the sample of males, females, 1^{st} graders, 2^{nd} graders, and 3^{rd} graders). Note that Cohen (1988) suggested that absolute value of *r* of .10 is classified as small, an absolute value of .30 is classified as medium and of .50 is classified as large.

CHAPTER 3

RESULTS

Descriptive Statistics

When the full sample of students (*N* = 140) was measured by the BRIEF2, on average, teachers reported the students global executive function problems were fairly low (*M* = 23.36) but there is evidence of considerable variability in problems across students (*SD* = 23.34). When the full sample of students (*N* = 140) were measured by the EYT tasks, the majority of students (M_G = 94.71%) accurately identified when they were to touch the fish in the "Go" task while students were able to accurately control their impulse in the "No-Go" task only 73.09% of the time. There is evidence of considerable variability in impulse control (SD_{NG} = 26.65). In the "Not This" task, considering there are at least 3 trails counted in each level of task, on average, students were able to complete 3 or 4 levels of the task measuring their phonological working memory (*M* = 11.84), which indicated that they can remember 3 or 4 characteristics corresponding to the auditory instruction. Likewise, considering there are at least 2 trails counted in each level of the task, on average, students in the "Mr. Ant" task were able to complete 3 or 4 levels of the task measuring their visual-spatial working memory (*M* = 7.52). In other words, children were able to remember the location of 3 or 4 colored stickers on the ant. In addition, on average, students were able to complete two sorting dimensions (i.e., color or shape) of the "Card Sorting" task and were able to complete a part of the more difficult border phase of the task (*M* = 14.12), which indicated their strong cognitive shifting skill. However, the variability was also considerable (*SD* = 3.39). For details of the descriptive statistics refers to Table 3.

Analyzing the data by sex (see Table 4), there are 45% female (*N* = 63), 55% male (N = 77). When the students' executive function problems (*N* = 140) were measured by BRIEF2, on average, the students' teachers reported females had fewer global executive function problems than males (M_m = 28.13, M_f = 17.54). Compared with females, there is also more variability in problems across males (SD_m = 25.62, SD_f = 18.80). In terms of the subscales such as inhibitory control, on average, the students' teachers reported females had much lower mean levels of inhibitory control problems than male (M_m = 4.60, M_f = 2.14) as well as less variability in problems than the male (SD_m = 5.05, SD_f = 2.88). However, there was not too much mean difference between females and males when students were measured by the EYT tasks. Even though in the cases of measuring shifting ("Card Sorting" task) and phonological working memory ("Not This" task), females performed a little better than male in shifting (M_m = 13.68, M_f = 14.64), but males got a little less problems in phonological working memory (M_m = 12.22, M_f = 11.38), the performance between male and female on these executive functioning tasks was similar. In sum, there were evident sex differences in teacher reports (i.e., BRIEF2) but not in direct assessments (i.e., EYT).

Analyzing the data by grade (see Table 5), there are 42.1% students (N = 59) in grade 1, 42.9% students (*N* = 60) in grade 2, and 15% students (*N* = 21) in grade 3. The results of both EYT and BRIEF2 indicated that the performance of children's executive functions became better with the increasing grade. For example, teacher reported GEC problems also decreased as students got older (M_1 = 29.15, M_2 = 20.63, M_3 = 14.90), although variability within grade was still quite high (SD_1 = 27.02, SD_2 = 18.88, SD_3 = 20.50). Likewise, the accuracy of inhibitory control (EYT "No-Go" task) in the direct assessments (EYT tasks)increased with the increasing grade (M_1 = 65.31, M_2 = 75.18, M_3 = 88.97), and variability within grade also decreased although it was still high (SD_1 = 28.84, SD_2 = 25.39, SD_3 = 12.90).

Correlations. Bivariate correlations among the three subscales (BRI, ERI, and CRI) in the BRIEF2 are presented in Table 6. The three subscales were significantly and strongly correlated with each other. Behavior regulation and emotion regulation were most strongly correlated, $r = .727$, $p < .001$, while cognitive regulation and emotion regulation were relatively weakly correlated, $r = .648$, $p < .001$. Bivariate correlations among the five EYT tasks are presented in Table 7. Three of the five tasks were significantly and moderately correlated with each other. Specifically, the correlation between phonological working memory ("Not This" task) and the behavioral skill ("Go" task) was not statistically significant ($r = .085$, $p = .321$), nor was the correlation between cognitive flexibility ("Card Sorting" task) and behavioral skills ("Go" task) ($r = .053$, $p = .53$). The correlations suggested that the different skills for the BRIEF2 may hang together (i.e., there is an amount of shared common variance that results in a factor of related items) in a single global score while the skills for the EYT may be more distinct.

Principal Components Analyses (PCAs) with a Fixed Number of Factors

The first research question—through the preliminary analysis, is it reasonable to assume that the expected constructs of teacher reports (three subscales versus one global scale) and direct assessments (two subscales versus one global scale) hold in a diverse sample of lower-level elementary school children—was addressed using a principal components analysis (PCA) for both the BRIEF2 and the EYT tasks first among the full sample of students (N = 140), and then among subgroups: males (N = 77) and females (N = 63) and 1st (N = 59), 2nd (N = 61), and 3rd (N = 21) grade. The PCAs were conducted to analyze several factors (broad regulatory domains) as well as one factor (global executive functions) in the BRIEF2 and EYT tasks for the full sample.

Given that I needed to further explore the factor structure and discuss whether one tool more adequately measures children's global executive functions than the other (i.e., my second research question), I considered only one factor (global executive functions) for each sex and grade after analyzing the full sample. Note that before conducting PCA on the BRIEF2, based on the instructions of the authors, it was necessary to remove the three items from the analysis: item 18, forgets his/her name; item 36, has trouble counting to three; and item 54, cannot find the front door if school because all students of the full sample or each group of sample (e.g., boys, girls, 1st grader, etc.) were rated by teachers with "N" or 0 point and SPSS cannot conduct PCA if a variable has zero variance. Therefore, there are currently 60 valid items in the BRIEF2.

The BRIEF2 for the full sample. To determine whether the 60 items from the BRIEF2 reflect three broad regulatory domains (i.e., behavior regulation, emotion regulation, and cognitive regulation) for the full sample, as suggested by the authors, I conducted a PCA with three fixed factors using 60 items. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the 60 items in the BRIEF2 (KMO = .883, χ^2 = 8988.474, p < .001). The eigenvalues for the first three components were 25.895, 5.889, and 3.220, respectively. In other words, the first factor accounted for more than 25 of the 60 units of variance (43.16%), while the second factor accounted for just under 6 units of variance (9.81%). Also, an additional eight components were revealed with eigenvalues over 1 (ranging from 2.470 to 1.018), which indicated that there could be up to eleven components extracted. This suggests that a three-component structure was not appropriate. However, there were some eigenvalues very close to 1 (e.g., 1.083, 1.068, and 1.018). If only following the rule of eigenvalue (greater than 1), it was statistically plausible to consider the BRIEF2 as an eleven-component structure. However, the fact that some eigenvalues were very close to 1 suggested that it might be plausible to consider the BRIEF2 as a structure with less than 11 components because theoretically there are not so many different domains under executive functions even though the narrow components tapped into more specific skills. Nevertheless, the number of subscales was relatively hard to determine because of the contradiction between statistics and our understanding of actual executive functions.

Even though the PCA with the item level data did not indicate a three-component structure, the authors of the BRIEF2 suggest that global executive function is comprised of behavior regulation, emotion regulation, and cognitive regulation. Given that the three-component structure is somehow reasonable based on the authors, I conducted a PCA with one fixed factor using the three composite scores (BRI, ERI, and CRI) to address the question of whether the BRIEF2 can be considered as a single global factor measuring lower-level elementary children's global executive function problems. In other words, I treated the three composite scores as items for running another PCA. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the three composite scores in the BRIEF2 (KMO = .730, $\chi^2 = 195.502$, $p < .001$). A single eigenvalue greater than 1 (2.352) was obtained. Figure 1 shows the factor loading of BRI is .898, the factor loading of ERI is .896, and the factor loading of CRI is .862, which are quite similar. The similar and large values of factor loadings suggested that the three regulatory domains were each contributed similarly and considerably to an overall measures of boys' global executive function problems. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a tool to adequately measure socio-demographically diverse lower-level elementary children's global executive function problems.

The EYT tasks for the full sample. To determine whether the five tasks from the EYT reflect two broad regulatory domains (i.e., cognitive regulation and behavior regulation) or only one global domain, I conducted a PCA with two fixed factors using the five scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the PCA was useful with the EYT data (KMO = .678, χ^2 = 58.418, p < .001). Two eigenvalues greater than 1 (1.885, 1.014) were obtained. Hence, based on the eigenvalue greater than 1 rule, there were two components in the EYT tasks (see Table 9) and the two-factor structure made sense for the full sample, which was also consistent with the original purposes of these tasks (i.e., the "Go/No-Go" tasks measure behavior regulation and the other three tasks measure cognitive regulation). Factor 1 (i.e., cognitive regulation) is comprised of the "Not This", "Mr. Ant", and "Card Sorting" tasks. Factor 2 (i.e., behavior regulation) is comprised of and the Go/No-Go task (including the accuracy of go and the accuracy of no-go). Figure 2 shows the factor loadings of five tasks with two components. However, one of the eigenvalues (1.014) was very close to 1. Hence, although the two-factor structure was appropriate for the full sample, it is the second factor accounts for less variance. Hence, the second factor might be a little less important to become a factor to measure diverse lower-level elementary children's global executive function skills. If considering the EYT tasks as a global tool, the factor loading of the "Go" task is .453, the factor loading of the "No-Go" task is .716, the factor loading of the "Not This" task is .639, the factor loading of the "Mr. Ant" task is .664, and the factor loading of the "Card Sorting" task is .564 (as Figure 3 shows). Although the factor loadings were not quite similar, all tasks contributed to the global executive functions but the "Go" task was less important than the others.

The BRIEF2 and the EYT tasks for the males. To address the question of whether the BRIEF2 can be considered as a single global factor to measure boy's global executive function problems, I conducted a PCA with one fixed factor using the three composite scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the three composite scores in the BRIEF2 (KMO = .727, χ^2 = 103.590, p < .001). A single eigenvalue greater than 1 (2.341) was obtained. Figure 4 shows the factor loading of BRI is .889, the factor loading of ERI is .900, and the factor loading of CRI is .860, which are quite similar. The similar and large values of factor loadings suggested that the three regulatory domains were each contributed similarly and considerable to an overall measures of children's global executive function problems. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a tool to measure socio-demographically diverse boy's global executive function problems.

For the EYT tasks, to address the question of whether the EYT tasks can be considered as a single global factor to measure boys' global executive functions, I also conducted a PCA with one fixed factor using the five scores from the four tasks. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the PCA is useful with the EYT data (KMO = .738, χ^2 = 44.462, p < .001). A single eigenvalue greater than 1 (2.096) was obtained. Figure 5 shows the factor loading of the "Go" task is .387, the factor loading of the "No-Go" task is .739, the factor loading of the "Not This" task is .711, the factor loading of the "Mr. Ant" task is .680, and the factor loading of the "Card Sorting" task is .658. Similar to the full sample, the "Go" tasks, as a part of task to measure children's behavior regulation, contributed weakly to an overall measure of children's global executive function problems than the other tasks. Based on these results, it appears that for the males the EYT tasks can be considered as a global tool to measure socio-demographically diverse boys' global executive function skills.

The BRIEF2 and the EYT tasks for the females. For the BRIEF2, to address the question of whether the BRIEF2 can be considered as a single global factor to measure girls' global executive function problems, I conducted a PCA with one fixed factor using the three composite scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the three composite scores in the BRIEF2 (KMO = .711, χ^2 = 97.488, $p < .001$). A single eigenvalue greater than 1 (2.387) was obtained. Figure 6 shows the factor loading of BRI is .919, the factor loading of ERI is .911, and the factor loading of CRI is .844, which are quite similar. The similar and large values of factor loadings suggested that the three regulatory domains were each contributed similarly and considerable to an overall measures of girls' global executive function problems. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a tool to measure socio-demographically diverse girls' global executive function problems.

For the EYT tasks, to address the question of whether the EYT tasks can be considered as a single global factor to measure girls' global executive functions, I also conducted a PCA with one fixed factor using the five scores from the four tasks. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the PCA was barely useful with the EYT data (KMO = .588, χ^2 = 23.284, p = .010). Two eigenvalues greater than 1 (1.668 and 1.270) were obtained (the detailed results of the second eigenvalue was actually hidden because of "forced" one factor), which indicated that there may be two factors in the EYT tasks for female. As a two-component structure tool, the "Go" task, "No-Go" task, and "Mr. Ant" task loaded to the factor 1 and the "Not This" task and "Card Sorting" task loaded to the factor 2. Figure 7 shows the factor loading of the "Go" task is .788, the factor loading of the "No-Go" task is .771, the factor loading of the "Mr. Ant" task is .608, the factor loading of the "Not This" task is .755, and the factor loading of the "Card Sorting" task is .798. Based on these results, it appears that for the females the EYT tasks cannot be considered as a global tool to measure socio-demographically diverse girls' global executive function skills.

The BRIEF2 and the EYT tasks for the first-grade students. For the BRIEF2, to address the question of whether the BRIEF2 can be considered as a single global factor to measure the first-grade students' global executive function problems, I conducted a PCA with one fixed factor using the three composite scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the three composite scores in the BRIEF2 (KMO = .730, χ^2 = 94.083, p < .001). A single eigenvalue greater than 1 (2.432) was obtained. Figure 8 shows the factor loading of BRI is .924, the factor loading of ERI is .891, and the factor loading of CRI is .885, which are quite similar. The similar and large values of factor loadings suggested that the three regulatory domains were each contributed similarly and considerable to an overall measures of the first-grade students' global executive function problems. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a single tool to measure the socio-demographically diverse first-grade students' global executive functions.

For the EYT tasks, to address the question of whether the EYT tasks can be considered as a single global factor to measure the first-grade students' global executive function skills, I also conducted a PCA with one fixed factor using the five scores from the four tasks. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the PCA was useful with the EYT data (KMO = .636, χ^2 = 22.362, p = .013). Two eigenvalues greater than 1 (1.781 and 1.145) were obtained, which indicated that there may be 2 factors in the EYT tasks for 1st graders. However, one of the eigenvalues (1.145) was close to 1. Therefore, although the two-factor structure was appropriate for the full sample, it is the second factor accounts for less variance. Hence, the second factor might be a little less important to become a factor to measure diverse lower-level elementary children's global executive function skills. Nonetheless, as a two-component structure tool, the "Go" task, "No-Go" task, and "Mr. Ant" task loaded to the factor 1 and the "Not This" task and "Card Sorting" task loaded to the factor 2. Figure 9 shows the factor loading of the "Go" task is .745, the factor loading of the "No-Go" task is .733, the factor loading of the "Mr. Ant" task is .672, the factor loading of the "Not This" task is .736, and the factor loading of the "Card Sorting" task is .801. Based on these results, it appears that for the males the EYT tasks cannot be considered as a global tool to measure the socio-demographically diverse first-grade students' global executive function skills.

The BRIEF2 and the EYT tasks for the second-grade students. For the BRIEF2, to address the question of whether the BRIEF2 can be considered as a single global factor to measure the second-grade students' global executive function problems, I conducted a PCA with one fixed factor using the three composite scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA was useful with the three composite scores in the BRIEF2 (KMO = .721, $\chi^2 = 66.306$, $p < .001$). A single eigenvalue greater than 1 (2.249) was obtained. Figure 10 shows the factor loading of BRI is .870, the factor loading of ERI is .879, and the factor loading of CRI is .849, which are quite similar. The similar and large values of factor loadings suggested that the three regulatory domains were each contributed similarly and considerable to an overall measures of the second-grade students' global executive function problems. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a single tool to measure the socio-demographically diverse second-grade students' global executive function problems.

For the EYT tasks, to address the question of whether the EYT tasks can be considered as a single global factor to measure the second-grade students' global executive function skills, I also conducted a PCA with one fixed factor using the five scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the PCA is useful with the EYT data (KMO = .658, χ^2 = 24.038, p = .008). Two eigenvalues greater than 1 (1.844 and 1.033) was obtained, which statistically indicated that there were 2 factors in the EYT tasks. However, the second eigenvalue (1.033) was very close to 1. Figure 11 shows the factor loadings of the tasks with two components. However, similar to applying the EYT tasks for the full sample, although the two-factor structure was appropriate for the full sample, it is still reasonable to consider the other factor might account for less variance. In other words, the second factor might be a little less important to be a factor to measure second-grade students' executive functions. If consider the EYT as a global tool, the factor loading of the "Go" task is .332, the factor loading of the "No-Go" task is .689, the factor loading of the "Not This" task is .671, the factor loading of the "Mr. Ant" task is .529, and the factor loading of the "Card Sorting" task is .728 (as Figure 12 shows). In this case, the "Go" task was less important to the global executive functions than the other tasks.

The BRIEF2 and the EYT tasks for the third-grade students. For the BRIEF2, to address the question of whether the BRIEF2 can be considered as a single global factor to measure the third-grade students' global executive function problems, I conducted a PCA with one fixed factor using the three composite scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the BRIEF2 indicated that the PCA is useful with the three composite scores in the BRIEF2 (KMO = .611, χ^2 = 24.656, p < .001). A single eigenvalue greater than 1 (2.214) was obtained. Figure 13 shows the factor loading of BRI is .866, the factor loading of ERI is .933, and the factor loading of CRI is .771, which are strong but not quite similar. In this case, the ERI was the most important domain to the global executive functions while the CRI was the least important domain. The strong factor loadings indicated that the three composite scores of regulations contributed strongly to an overall measure of children's global executive function problems in this case. Based on these results, it appears that for the full sample the BRIEF2 can be considered as a tool to measure the socio-demographically diverse third-grade students' global executive functions.

For the EYT tasks, to address the question of whether the EYT tasks can be considered as a single global factor to measure the third-grade students' global executive functions, I also tried to conduct a PCA with one fixed factor using the five scores. The results of Kaiser-Meyer-Olkin and Bartlett's test in the EYT tasks indicated that the data was pretty poor to conduct PCA (KMO = .593, χ^2 = 11.399, p = .327). Hence, the PCA for the EYT tasks as a tool to measure the global executive functions for the third-grade students cannot be conducted.

Correlation Analyses

Based on the factor loadings weight of the composite scores for BRI, ERI, and CRI in the BRIEF2 and the scores in the EYT tasks, a weighted average reflecting global executive functioning was created for each assessment. One thing needed to be acknowledged was that I created the single composite scores despite evidence suggesting the EYT might be better represented by two components. And the reason why I created the single composite score was that all cases the eigenvalue of the second factor was close to 1 so the second factor might be not as meaningful. For the purpose of illustrating whether teacher reports and direct assessments of children's global executive functions are correlated, bivariate correlations among the total score of the BRIEF2 and the EYT tasks in each group are presented in Table 8.

Given that the BRIEF2 is a tool to measure children's executive function problems while the EYT tasks is to measure children's executive function skills, the correlation between them has to be negative. The more problems reflected by the BRIEF2, the poorer skills the EYT tasks should indicate. Using the full sample (N = 140), the global executive function score from the BRIEF2 was significantly and moderately correlated with the global executive function scores from the EYT tasks, r = -.38, p < .001. Noted that the "significantly" means that the association is predicted not zero in the population, and "moderately" refers to the effect size to describe the degree of correlation, based on Cohen's suggestion. For males only (N = 77), the global executive function score from the BRIEF2 was significantly and moderately correlated with the global executive function score from the EYT tasks, r = -.405, p < .001. For females only (N = 63), the global executive function score from the BRIEF2 was significantly and moderately correlated with the global executive function score from the EYT tasks, r = -.37, p = .003. So, the correlation between the global executive function score from the BRIEF2 and the global executive function score from the EYT tasks across the sample of males was a little stronger than across the sample of females. For the first-grade students only (N = 59), the global executive function score from the BRIEF2 was significantly and moderately correlated with the global executive function score from the EYT tasks, r = -.346, p = .007. For the second-grade students only (N = 60), the global executive function score from the BRIEF2 was significantly and moderately correlated with the global executive function score from the EYT tasks, r = -.386, p = .002. However, the global executive function score from the BRIEF2 was not significantly correlated with the global executive function score from the EYT tasks across the sample of third-grade students (N = 21), r = -.024, p = .917. Hence, the correlation for the sample of the second-grade students was the strongest among all three grades students.

CHAPTER 4

DISCUSSION

With growing interest in both the antecedents and consequences of executive functions, there is a need to identify the best and most efficient approaches to measuring these skills. In the current study, I used data collected for the *LCDSLS* project to examine the preliminary factorial structure of a teacher report and direct assessment tool assessing executive functions in a sample of diverse elementary school children in the mid-west and began to offer insight into potential gender and grade differences in these structures. Ultimately, I sought to lend insight into which tool can more adequately measure children's global executive functions for researchers and practitioners who are interested in these skills.

The findings provided some evidence that teacher reports as a global tool adequately measured diverse lower-elementary school children's executive functions while the structure with three subscales (i.e., cognitive, behavioral, and emotion regulation) of the teacher reports did not adequately reflect their executive functions in this sample. In fact, analyses using the eigenvalue greater than 1 rule revealed that an eleven-component structure might be more valid than the three-component structure. The validity of the measures may be affected by the complexity of the structure. In other words, it is possible that the different subscales of teacher reports measured children's executive functions differently. Further, some subscales of the teacher reports could be separated into more distinct categories, as the findings suggest, and subscales titles could not reflect the subscales items enough (Duku & Vaillancourt, 2014; Spiegel et al., 2017). For example, in children this age, the broad skill of cognitive regulation may be better separated into some narrower subscales including initiating tasks, working memory, planning, task-monitoring, and organization of materials (Gioia et al, 2015). Hongwanishkul, Happaney, Lee, and Zelazo (2005) suggested that cognitive and behavioral regulation (i.e., hot and cool regulation) may not be two distinct sets of skills until children are older. Therefore, perhaps attempting to measure broad skills like cognitive regulation, behavioral regulation, and emotion regulation is not appropriate until these skills begin to differentiate at older ages. Instead either very narrow subscales tapping into specific skills or a global skill might better reflect where children this age are developmentally.

The findings also suggested that direct assessments can be used as a tool with two subscales including cognitive regulation and behavior regulation to measure diverse lower-elementary school children's executive functions in this full sample. The second component (i.e., behavior regulation), however, might be a little less important to measure their executive functions because the eigenvalue of the second factor was very close to 1. In other words, the findings of the direct assessments indicated that the second factor needed to be interpreted with caution. One of the limitations of direct assessments—task impurity problem (Miyake et al., 2000; Toplak et al., 2013)—may be a reason for this dilemma. Developers of teacher reports can develop items as unidimensional as possible. However, as Miyake et al. (2000) suggested, one direct assessment task can focus on more than one executive functioning skill, which may be reflect the multidimensionality of the skill. For example, in terms of the Go/No-Go tasks, the tasks were designed to focus on children's behavior regulation, but children still needed to have basic cognitive skills such as working memory to complete the task.

In terms of measuring the diverse lower-level elementary students' executive functions in the full sample, the reasons why the teacher report fit better as a global tool than the direct assessments include the development of children's executive functions in this age and the characteristics of the tool itself. From the perspective of children's development, it is hard for lower-level elementary children (6-9 ages of the year) to differentiate the interactive set of skills such as cognitive regulation, behavior regulation and emotion regulation under the umbrella of executive functions. The boundaries among each set of skills are not clear for children at this age. Indeed, Zelazo and Cunningham (2007) suggested that in highly emotional situations, emotion regulation may actually interfere with cognitive regulation but in less emotional situations, emotion regulation may support cognitive regulation (or take a back seat to it while children draw on their cognitive regulation skills). Hence, the skills may not be as independent or distinguishable as would be ideal for measuring them. Further, children develop their emotion regulation skills when they get older. Hence, children at this age may require global tools to assess their broad set of executive functions skills rather than tools that assess specific skills. Therefore, the characteristics of the tools are also important. Relatedly, the 63 items in the teacher report were so comprehensive that they can almost cover all dimensions of executive functions, which means the teacher report can be used as a global tool but each individual scale may not have had sufficient items to reflect that skill at this developmental stage. In contrast, there are only four tasks to measure cognitive regulation and behavior regulation, which is not enough to cover all domains of children's executive functions, compared with the seven skills in the teacher report measuring these two constructs. So, the direct assessments may not measure children's global executive functions as adequately as the teacher report because of the small number of tasks. Nonetheless, there is the issue of task impurity with the direct assessments (Miyake et al., 2000; Toplak et al., 2013), which enables the direct assessments to be used to measure global executive functions to some extent.

Preliminary Findings by Sex

When measuring the boys' and girls' global executive functions, the teacher reports, as a global tool, appear to more adequately measure children's executive functions than the direct assessment. Even though teacher reports measured adequately as a global tool because there was no obvious difference of factor structure between boys and girls, in fact, the descriptive statistics revealed that the scores look different for boys and for girls. That is, the mean number of boys' executive function problems was larger than the mean number of girls' executive function problems, which suggested that teachers believed boys performed worse than girls in the classroom. This potential difference in the score may be due to teachers' bias. Teacher reports are biased with students' executive functions (Garcia, et al., 2019; Nisbett & Wilson, 1977; Ready & Wright, 2011). There are reasons for this bias on sex. Because boys may perform more active than girls of the same ages and it is hard for teachers to manage boys, a negative perception or bias comes to the teachers' mind. Moreover, compared with children's cognitive regulation, it is easier for teachers to rate the children's behavior regulation. Given that boys' behavior regulation seems poorer than girls', teachers may believe that the number of boys' executive function problems is larger even though behavior regulation is only one of the dimensions of executive functions. Despite these mean differences, preliminary findings suggested that the factor structure is similar and thus additional work is needed to more adequately understand these mean level differences and to determine whether the factor structure is, in fact, similar for boys and girls.

Interestingly, according to the children's performance on the direct assessments, there was not too much difference in the mean level scores between boys and girls, which suggested that the direct assessments may offer a way to avoid the issue of teachers' bias. However, the preliminary subgroup analysis by sex suggested that factor structure may look different for boys and for girls. That is, for boys there appears to be one factor while for girls, two factors. Compared with boys, girls might develop their behavior regulation and cognitive regulation earlier. For example, there is initial evidence that boys (5-11 years) make more errors and have worse accuracy than girls in this age in attention functions like inhibitory control (Sobeh and Spijkers, 2012). Boys in this age might not yet differentiate the behavior regulation and cognitive regulation as much as girls. Further, girls might use all of their inter-related skills while boys might rely more heavily on one skill so it appears that executive function is more unidimensional for boys. Thus, it might be more appropriate to measure girls' executive functions with the two-factor direct assessments and to measure boys' executive functions with the single global teacher report. Therefore, when using a teacher report, boys and girls can both be described by a single global executive function score but when using direct assessments, it may be better to treat girls' executive functions as multi-dimensional and boys' executive functions as unidimensional. Additional research based on different samples should be conducted to continue to explore this question.

Preliminary Findings by Grade

When it comes to the issue of using teacher reports and direct assessment to measure children's executive functions in different grades, the findings of descriptive statistics revealed mean differences between grades and suggested that the performance of children's executive functions improved with the increasing grade, which reflected the typical development of children's executive functions. Executive functions become more refined as children get older. This would be expected as both children's behavior and cognitive regulation develop rapidly in the elementary school-age period (Carlson, 2005; Diamond, 1990; Garon, et al., 2008; Hughes, et al., 2010; O'Meagher et al., 2019).

The preliminary findings suggested that the teacher report as a single global tool measured children's global executive functions adequately in each grade. However, the finding suggested that it might be better to use the direct assessment as a two-factor tool to measure children's executive functions in first and second grade (the results with the third grader sample cannot be presented because of the small sample size, but I can assume that the third graders also fit two-factor direct assessment). Admittedly, the second factor might be a little less important to measure children's executive functions because the eigenvalues in both cases are close to 1. Given that children in each subgroup of grade also included both males and females, the sample of each subgroup of grade can be considered as similar as a "smaller" full sample. Hence, similar to the full sample, as mentioned above, task impurity and the young age might explain this situation. Direct assessments can focus on more than one executive functioning skill and the skills may not be distinguishable for young children, which may enable the direct assessments to be used to assess global executive functions to some extent. So, the second factor of executive functions needed to be interpreted with caution (details refer to page 37 and 38 in the paper).

Correlations between Teacher Reports and Direct Assessments

Consistent with prior research (Fuhs et al., 2015) which reported that several teacher reports and direct assessments of executive functions in their sample of prekindergarten children were correlated (the range of r is from .257 to .511), the findings in the current study provided the evidence that the teacher report and the direct assessment tool are significantly and moderately correlated with each other in the full sample of socio-demographically diverse lower-level elementary school children (r = -.38). Fuhs et al. (2015) also suggested that the two measures of executive functions were correlated especially when the direct assessments were modeled as a single component score. Similarly, the findings in the current study suggested the moderate correlation between the two measures of global executive functions for the full sample.

However, the correlations summarized in the current study and in the study by Fuhs et al. (2015) were stronger than other those summarized by Blair et al. (2015), Clark et al. (2010), Mcauley et al, (2010), O'Meagher et al. (2019), Sulik et al. (2010), and Toplak et al. (2013). The reason for the difference might be the various demographic characteristics and the various ages of the sample and in research. For example, Blair et al. (2015) used a primarily rural and low-income sample of prekindergarten through second grade children. Low-income children are already behind their more advantaged peers in typical learning skills like reading and math due to a lack of education supports from family or to excessive stress (Evans, 2004). The children from low-income families or from regions of high poverty may have little chance to use iPad before, so it may be hard for them to proficiently use iPad to participate the direct assessments. Hence, the weak correlation between the teacher report and the direct assessments may be explained by this demographic characteristic. Furthermore, the moderate correlation in my sample indicated that to some extent the two measures have similar functions and can tap into similar broad skills for younger children. That does not indicate that teacher reports and direct assessments are the same, however; that means the difference between the two measures are not clear because younger children have poorer executive functions and they cannot distinguish different dimensions in the executive function set. As Toplak et al. (2013) suggested, direct assessments focus on information-processing mechanisms in brain (e.g., working memory and inhibition) while teacher reports focus more on the choice of action to achieve goals and decision-making. Overall, they may be related functions, which is why they correlated but they are distinct. It is really hard for younger children to differentiate these executive function skills. Because the sample used in the research by Fuhs et al. (2015) was prekindergarten children and the sample used in the current study were lower-level elementary school children, both the samples might be too young to distinguish the different dimensions of executive functions. Therefore, the difference between the teacher report and the direct assessment was not clear when measuring young children like the sample in the current study, which may explain the limited relative moderate correlations.

In terms of subgroups (males, females, 1st graders, 2nd graders, and 3rd graders), one thing that needs to be noted was that I used the single composite scores for all subgroups despite evidence suggesting the direct assessments might be better represented by two components in some subgroups, because all cases the eigenvalue of the second factor was close to 1. So, the second factor might be not as meaningful. Nevertheless, these findings provide preliminary evidence that even in the subgroups, a global score for executive functions for the teacher report and the direct assessment are significantly and moderately correlated with each other (the range of r is from -.405 to -.346) except for third-grade children (r = -.024). The reasons why the correlation in the sample of third-grade children was weak may include the elder age of third-grade children and the small sample size. Given that the sample of the research by Fuhs et al. (2015) was prekindergarten children and the sample of the current study which showed moderate correlation was first and second-grade elementary school children, and the samples of other research were diverse with a wide range of ages or even elder sample, I can predict that the correlation between teacher reports and direct assessments is associated with the age of the sample. The older the children become, the more developed executive function skills the children have, the weaker correlation between teacher reports and direct assessments may be obtained.

Limitations and Future Research

Although the results from the current study contribute to our knowledge about the structure of teacher reports and direct assessments of children's executive functions and provide some insight into the relationship between teacher reports and direct assessments, especially digging into their structures, study limitations must be acknowledged. First, the data analysis is preliminary, and compared with confirmatory factor analysis within a structural equation modeling framework, PCA may not be the best method to use. In factor analysis, total variance can be partitioned into common variance and unique variance (unique variance is equal to the sum of specific variance and error variance) (UCLA: Statistical Consulting Group). However, PCA is not a model and makes the assumption that there is no unique variance and the total variance is equal to common variance, hence, no separation is made between common variance and unique variance and the initial communalities are all 1, which represents all of the variance of each item included in the analysis (Steyn, 2017). And PCA makes no assumption about a model and focuses on which linear relationships exist and how any particular variable might contribute to that relationship (Steyn, 2017).

Second, PCA may not be the best method to use, particularly if the measurement model is group invariant. Given that I also focused on whether the structure of the measures held in subgroups, the samples for each analysis were technically different as I conducted separate PCAs for each subgroup; hence, the groups were not necessarily comparable. In other words, within the full sample (N = 140), I selected five different samples with only males (N = 77), only females (N = 63), only 1st graders (N = 59), only 2nd graders (N = 60), and only 3rd graders (N = 21) to explore if the structure held in each subgroup. Actually, these samples were mutually exclusive groups as were the grades but that males were distributed across grades (and vice versa). In other words, they were not five independent samples; they were two independent samples for the sex subgroup analysis and three independent samples for the grade subgroup analyses. The sample and the sample size were both various. However, if I conducted confirmatory factor analysis within a structural equation modeling framework, the sample would always be the full sample ($N = 140$), which would allow me to more accurately examine measurement invariance across groups than conducting PCAs. I currently do not have enough skills to conducting confirmatory factor analysis within a structural equation modeling framework, but for future research, more advanced data analysis methods are needed.

Third, in the first PCA I conducted to explore if the three-component structure is appropriate, the results showed that the three-component structure is not adequately to measure children's executive functions. However, in the second PCA I conducted to explore if the single global structure can be considered as a valid structure (i.e., consider the three broadband items into one component), I still used the three composite scores as items to fit the global executive composite based on the authors' suggestion, which should be acknowledged as one of the limitations in the research. However, the authors also suggest that there are 9 subscales that can be derived from this measure and the item level data (i.e., the results of the first PCA) better reflect this. Moreover, it is possible that for children at this age, the three broad subscales have not yet differentiated and thus are not reflected in the first PCA so these analyses should be interpreted with some caution.

Last but not least, the sample size for third-grade children was small, which made it difficult to assess the results for the sample of third-grade children. For better conducting confirmatory factor analysis within a structural equation modeling framework, more than 200 participants might be ideal.

Implications

Based on the findings and analysis, I made suggestions to the researchers who are interested in measuring diverse lower-level elementary school children's executive functions. First, if researchers are more interested in measuring children's global executive functions and have limited time and budget to conduct both measures, they can use only the teacher reports (e.g., different versions of the BRIEF) rather than direct assessments, because the findings suggested that the teacher reports can adequately measure the children's global executive functions as a single global tool. Second, if researchers intend to measure boys' global executive functions, both teacher reports and direct assessments can do so adequately. If researchers intend to measure girls' executive functions, I would suggest using direct assessments (e.g., EYT) with subscales and focus on two dimensions (e.g., cognitive regulation and behavior regulation) in executive functions, which is not only because direct assessments might be good at measuring two dimensions, but also because girls in this age may have better skills to distinguish the specific executive function skills. Third, if researchers want to avoid teachers' bias while they have enough time and budget to conduct direct assessments, they would better to use both. Last but not least, if researchers intend to measure older children or children with relatively stronger executive functions, I suggested using both measures because the greater inconsistency between teacher reports and direct assessments might exist as children grow. However, as previously mentioned, all the suggestions are preliminary and may not be relevant for a different sample, and researchers need to conduct additional exploratory research like this if they are interested in different population.

References

Alloway, T.P., Gathercole, S.E., & Kirkwood, H.J. (2008). *Working Memory Rating Scale*. London: Pearson Assessment.

Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71.

Anna B. Costello, & Jason Osborne. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *7*, 1.

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum Associates. *Inc. Anderson The adaptive character of thought1990*.

Baddeley, A. D. (1986). Working memory. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 8, pp. 47–90). New York: Academic Press.

Baker, C. N., Tichovolsky, M. H., Kupersmidt, J. B., Voegler-Lee, M. E., & Arnold, D. H. (2015). Teacher (mis) perceptions of preschoolers' academic skills: Predictors and associations with longitudinal outcomes. *Journal of educational psychology*, *107*(3), 805.

Bari, A., & Robbins, T. (2013). Noradrenergic versus dopaminergic modulation of impulsivity, attention and monitoring behavior in rats performing the stop-signal task. *Psychopharmacology*, *230*(1), 89–111.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational intelligence*, *4*(3), 349-355.

Bernier, A., Matte-Gagné, C., & Bouvette-Turcot, A. A. (2014). Examining the interface of children's sleep, executive functioning, and caregiving relationships: A plea against silos in the study of biology, cognition, and relationships. *Current Directions in Psychological Science*, *23*(4), 284-289.

Blair, C., & Ursache, A. (2011). A bidirectional model of executive functions and self-regulation.

Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter–word knowledge in the early elementary grades. *Developmental psychology*, *51*(4), 459.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development*, *78*(2), 647-663.

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

Borella, E., Carretti, B., & Pelegrina, S. (2010). The Specific Role of Inhibition in Reading Comprehension in Good and Poor Comprehenders. *Journal of Learning Disabilities*, *43*(6), 541–552.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational intelligence*, *4*(3), 349-355.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337-349.

Brooks, P. J., Hanauer, J. B., Padowska, B., & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development*, *18*(2), 195-215.

Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*(4), 263-272.

Canas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, *46*(5), 482-501.

Carlson, S. M. (2005). Developmentally Sensitive Measures of Executive Function in Preschool Children. *Developmental Neuropsychology*, *28*(2), 595–616.

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York, NY: Academic Press.

Clark, C. A., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental psychology*, *46*(5), 1176.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Academic press.

Conway A, Stifter CA, Conway, A., & Stifter, C. A. (2012). Longitudinal antecedents of executive function in preschoolers. *Child Development*, *83*(3), 1022–1036.

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk in kindergarten. *Early Childhood Research Quarterly, 3,* 1–19. doi:10.1016/0885- 2006(88)90026-9

Cooper, D. H., & Farran, D. C. (1991). *The Cooper-Farran Behavioral Rating Scales*. Brandon, VT: Clinical Psychology.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain* (Doctoral dissertation, ProQuest Information & Learning).

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

Dennett, D. C. (1989). *The intentional stance*. MIT press.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching a. *Annals of the New York Academy of Sciences*, *608*(1), 637-676.

Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, *21*(5), 335–341.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental psychobiology*, *29*(4), 315-334.

Duku, E., & Vaillancourt, T. (2014). Validation of the BRIEF-P in a sample of Canadian preschool children. Child Neuropsychology, 20(3), 358–371.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2008). "School readiness and later achievement": Correction to Duncan et al (2007). *Developmental Psychology*, *44*(1), 232. https://doi-org.proxy-remote.galib.uga.edu/10.1037/0012-1649.44.1.217

Evans, G. W. (2004). The environment of childhood poverty. *American psychologist*, *59*(2), 77.

Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educational Studies in Mathematics*, *21*(1), 55-69.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101-135.

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. Developmental Psychology, 47(2), 404–416.

Fuhs, M. W., Farran, D. C., & Nesbitt, K. T. (2015). Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. *Journal of Educational Psychology*, *107*(1), 207.

Garcia, E. B., Sulik, M. J., & Obradović, J. (2019). Teachers' Perceptions of Students' Executive Functions: Disparities by Gender, Ethnicity, and ELL Status. *Journal of Educational Psychology*, (5), 918.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin*, *134*(1), 31.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(1), 1-16.

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). Behavior Rating Inventory of Executive Function-Preschool Version. Retrieved from http://search.ebscohost.com.proxy-remote.galib.uga.edu/login.aspx?direct=true&db=loh&AN=test.2722&site=eds-live

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior rating inventory of* *executive function: BRIEF*. Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). Behavior Rating Inventory of Executive Function, Second Edition. Retrieved from http://search.ebscohost.com.proxy-remote.galib.uga.edu/login.aspx?direct=true&db=loh&AN=test.8651&site=eds-live

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology*, *8*(4), 249-257.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Psychology press.

Hammond, S. I., Carpendale, J. I. M., Bibok, M. B., Müller, U., & Liebermann-Finestone, D. P. (2012). The Effects of Parental Scaffolding on Preschoolers' Executive Function. *Developmental Psychology*, *48*(1), 271–281. https://doi-org.proxy-remote.galib.uga.edu/10.1037/00025519

Hand, S., Rice, L., & Greenlee, E. (2017). Exploring teachers' and students' gender role bias and students' confidence in STEM fields. *Social Psychology of Education*, *20*(4), 929-945.

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtis, G. (1993). Wisconsin Card Sorting Test (WCST) manual, revised and expanded. Odessa, FL: Psychological Assessment Resources.

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of educational psychology*, *101*(3), 662.

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

Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, *31*, 1-21.

Howard, S. J., & Melhuish, E. (2017). An early years toolbox for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. *Journal of Psychoeducational Assessment*, *35*(3), 255-275.

Howard, S. J., & Okely, A. D. (2015). Catching fish and avoiding sharks: Investigating factors that influence developmentally appropriate measurement of preschoolers' inhibitory control. *Journal of Psychoeducational Assessment*, *33*(6), 585-596.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British journal of developmental psychology*, *16*(2), 233-253.

Hughes, C., Ensor, R., Wilson, A., and Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. Developmental Neuropsychology, 35(1), 20–36.

Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of school psychology*, *43*(4), 303-320.

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. *Child Development*, *77*, 1822-1841.

Jensen, A. R., & Rohwer Jr, W. D. (1966). The Stroop color-word test: a review. *Acta psychologica*, *25*, 36-93.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational & Psychological Measurement*, *20*(1), 141.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, *34*(1), 111-117.

Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of personality*, *71*(6), 1087-1112.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59-80

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*(2), 163.

Man, D., & Vision, A. (1982). A computational investigation into the human representation and processing of visual information

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

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the Behavior Rating Inventory of Executive Function more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society*, *16*(3), 495–505.

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental psychology*, *43*(4), 947.

McKinnon, R. D., & Blair, C. (2018). Does early executive function predict teacher–child relationships from kindergarten to second grade?. *Developmental psychology*, *54*(11), 2053.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology, 46,* 235–261.

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), 8-14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.

Morra, S. (1994). Issues in working memory measurement: Testing for M capacity. *International Journal of Behavioral Development*, *17*, 143-159.

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school.

National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB). (2013). *Monographs of the Society for Research in Child Development*, *78*(4), 1-172.

Newell, A. (1982). The knowledge level. *Artificial intelligence*, *18*(1), 87-127.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology, 35,* 250 –256.

Obradović, J., Sulik, M. J., Finch, J., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology, 55,* 4–13.

O'Meagher, S., Norris, K., Kemp, N., & Anderson, P. (2019). Examining the relationship between performance-based and questionnaire assessments of executive function in young preterm children: Implications for clinical practice. *Child Neuropsychology*, *25*(7), 899-913.

Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task?. *Infant and Child Development: An International Journal of Research and Practice*, *11*(2), 93-105.

Pollock, J. L. (1995). *Cognitive carpentry: A blueprint for how to build a person*. Mit Press.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental psychology*, *45*(3), 605.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal*, *48*(2), 335-360.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental psychology*, *45*(4), 958.

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

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of experimental psychology: General*, *132*(4), 566.

Scarupa, H. J. (2014). Measuring Elementary School Students' Social and Emotional Skills: Providing Educators with Tools to Measure and Monitor Social and Emotional Skills That Lead to Academic Success. Publication# 2014-37. *Child Trends*.

Shettle, C., Roey, S., Mordica, J., Perkins, R., Nord, C., Teodorovic, J., ... & Brown, J. (2007). The Nation's Report Card [TM]: America's High School Graduates. NCES 2007-467. *National Center for Education Statistics*.

Sobeh, J., & Spijkers, W. (2012). Development of attention functions in 5-to 11-year-old Arab children as measured by the German Test Battery of Attention Performance (KITAP): A pilot study from Syria. *Child Neuropsychology*, *18*(2), 144-167.

Spiegel, J. A., Lonigan, C. J., & Phillips, B. M. (2017). Factor structure and utility of the behavior rating inventory of executive function-preschool version. Psychological Assessment, 29(2), 172–185.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Steyn, P. (2017, April 4). Which test: factor Analysis (FA, EFA, PCA, CFA) [Web log post]. Retrieved from https://www.introspective-mode.org/factor-analysis-fa-efa-pca-cfa

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643–662.

Sulik, M. J., Huerta, S., Zerr, A. A., Eisenberg, N., Spinrad, T. L., Valiente, C., ... & Edwards, A. (2010). The factor structure of effortful control and measurement invariance across ethnicity and sex in a high-risk sample. *Journal of psychopathology and behavioral assessment*, *32*(1), 8-22.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.

Thorell, L.B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*, 33, 536–552.

Thorell, L.B., Eninger, L., Brocki, K.C., & Bohlin, G. (2010). Childhood Executive Function Inventory (CHEXI): A promising measure for identifying young children with ADHD? *Journal of Clinical and Experimental Neuropsychology*, 32, 38–43.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct?. *Journal of Child Psychology and Psychiatry*, *54*(2), 131-143.

UCLA: Statistical Consulting Group. Principal components (PCA) and exploratory factor analysis (EFA) with SPSS. (n.d.). Retrieved from https://stats.idre.ucla.edu/spss/seminars/efa-spss/#s2

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self- regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives, 6,* 122–128.

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

Wilson, B.A., Alderman, N., Burgess, P.W., Emslie, H., & Evans, J.J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. UK: Thames Valley Test Company.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297-301.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs for the Society for Research in Child Development*, *78*(4), 16-33.

Zelazo, P. D., & Cunningham, W. A. (2007). Executive Function: Mechanisms Underlying Emotion Regulation.

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child development perspectives*, *6*(2), 112-121.

Appendix A

Tables

Table 1

List of Teacher Reports of Children's Executive Functions and the Relevant Cognitive Abilities

Teacher reports	Factors or subscales
Cooper-Farran Behavioral Rating Scale (CFBRS; Cooper & Farran, 1988, 1991)	Two subscales: Work-Related Skills (WRS) and Interpersonal Skills (IPS)
Behavioral Assessment of the Dysexecutive Syndrome (BADS; Wilson et al., 1996)	Single Scale designed to measure emotional/personality, motivational, behavioral, and cognitive changes
Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000)	Inhibit, Shift, and Emotional Control scales form the Behavioral Regulation Index (BRI); Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor scales form the Metacognition Index (MI); Behavioral Regulation Index and Metacognition Index form a Global Executive Composite (GEC)
Children's Behavior Questionnaire (Rothbart, Ahadi, Hershey, & Fisher, 2001)	Three broad dimensions of temperament: Extraversion/Surgency, Negative Affectivity, and Effortful Control
Behavior Rating Inventory of Executive Function, Preschool Version (BRIEF-P; Gioia, Espy, & Isquith, 2003)	The Inhibitory Self-Control Index (ISCI)is composed of the Inhibit and Emotional Control scales, the Flexibility Index (FI) is composed of the Shift and Emotional Control scales, and the Emergent Metacognition Index (EMI) is composed of the Working Memory and Plan/Organize scales.
Working Memory Rating Scale (WMRS; Alloway et al., 2008)	Single composite measure of working memory deficits
Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008; Thorell et al., 2010)	Working memory and inhibition

Teacher Survey in Measuring Elementary School Students' Social and Emotional Skills (Scarupa, 2014)	Self-control, persistence, and social competence.
Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2; Gioia, Isquith, Guy, & Kenworthy, 2015)	Inhibit and Self-monitor scales form the Behavioral Regulation Index (BRI); Shift and Emotional Control scales form the Emotional Regulation Index (ERI); and Initiate, Working Memory, Plan/Organize, Task-monitor, Organization of Materials scales form the Cognitive Regulation Index (CRI). The three indexes form a Global Executive Composite (GEC)

Table 2

List of Some Direct Assessments of Executive Functions

Dimension	Direct Assessment	Author (Year)
Working Memory	Corsi Blocks Task—Visuo-spatial short-term memory and working memory	Corsi, 1972
	Mr. Cucumber Task	Case, 1985
	EYT Mr. Ant Task—Visual-spatial working memory	Howard and Melhuish, 2017
	EYT Not This Task—Phonological working memory	Howard and Melhuish, 2017
Inhibitory Control	Peg Tapping (PT)	Diamond and Taylor, 1996
	Head Toes Knees Shoulders (HTKS)	Ponitz, McClelland, Matthews, and Morrison, 2009
	EYT Go/No-Go Task	Howard and Melhuish, 2017
Cognitive Flexibility (Shifting)	Stroop Test	Stroop, 1935
	Wisconsin-Card Sorting Test (WCST)	Chelune, Talley, Kay, and Curtis, 1993
	The Brixton Spatial Anticipation Test	Burgess and Shallice, 1996
	Dimensional Change Card Sort (DCCS)	Zelazo, 2006
	EYT Card Sorting Task	Howard & Melhuish, 2017

Table 3

Descriptive Statistics for the Full Sample

	Variable	Full Sample Mean/% (SD)
EYT (Direct Assessment)	GO-ACC	94.71% (6.80)
	NO-GO-ACC	73.09% (26.65)
	NOT THIS-ACC	11.84 (3.96)
	MR.ANT-ACC	7.52 (3.22)
	CSORT-ACC	14.12 (3.39)
BRIEF2 (Teacher Report)	Inhibit	3.49 (4.38)
	Self-Monitor	1.91 (2.59)
	BRI	5.41 (6.76)
	Shift	2.36 (2.73)
	Emotional Control	1.96 (3.34)
	ERI	4.33 (5.61)
	Initiate	1.79 (2.24)
	Working Memory	4.04 (4.28)
	Plan/Organize	3.03 (3.69)
	Task-Monitor	3.60 (3.25)
	Organization of Materials	1.17 (1.87)
	CRI	13.63 (13.80)
	GEC	23.36 (23.34)

Table 4

Descriptive Statistics for the Males and Females (by Sex)

	Variable	Male Mean/% (SD)	Female Mean/% (SD)
EYT (Direct Assessment)	GO-ACC	95.22% (6.92)	94.09% (6.66)
	NO-GO-ACC	72.66% (27.17)	73.62% (26.21)
	NOT THIS-ACC	12.22 (4.05)	11.38 (3.83)
	MR.ANT-ACC	7.61 (3.12)	7.41 (3.36)
	CSORT-ACC	13.68 (3.68)	14.67 (2.92)
BRIEF2 (Teacher Report)	Inhibit	4.60 (5.05)	2.14 (2.88)
	Self-Monitor	2.36 (3.00)	1.37 (1.85)
	BRI	6.96 (7.85)	3.51 (4.50)
	Shift	2.53 (2.81)	2.16 (2.64)
	Emotional Control	2.27 (3.61)	1.59 (2.97)
	ERI	4.81 (5.94)	3.75 (5.16)
	Initiate	2.13 (2.41)	1.37 (1.96)
	Working Memory	4.83 (4.76)	3.08 (3.40)
	Plan/Organize	3.52 (4.15)	2.43 (2.96)
	Task-Monitor	4.40 (3.35)	2.62 (2.87)
	Organization of Materials	1.48 (2.03)	.79 (1.59)
	CRI	16.36 (15.02)	10.29 (11.37)
	GEC	28.13 (25.62)	17.54 (18.80)

Table 5

Descriptive Statistics for the 1st, 2nd, and 3rd Grade Students (by Grade)

	Variable	Grade 1 Mean/% (SD)	Grade 2 Mean/% (SD)	Grade 3 Mean/% (SD)
EYT (Direct Assessment)	GO-ACC	94.03% (5.89)	94.27% (8.28)	97.91% (2.46)
	NO-GO-ACC	65.31% (28.84)	75.18% (25.39)	88.97% (12.90)
	NOT THIS-ACC	11.31 (3.14)	12.08 (4.52)	12.67 (4.29)
	MR.ANT-ACC	6.17 (2.90)	8.22 (3.08)	9.33 (3.07)
	CSORT-ACC	13.17 (3.72)	14.78 (3.05)	14.90 (2.70)
BRIEF2 (Teacher Report)	Inhibit	4.58 (4.56)	2.73 (3.88)	2.62 (4.74)
	Self-Monitor	2.31 (2.65)	1.72 (2.46)	1.38 (2.77)
	BRI	6.88 (6.95)	4.45 (6.15)	4.00 (7.40)
	Shift	2.63 (3.07)	2.30 (2.43)	1.81 (2.56)
	Emotional Control	2.81 (4.02)	1.32 (2.43)	1.43 (3.12)
	ERI	5.44 (6.68)	3.62 (4.27)	3.24 (5.38)
	Initiate	2.25 (2.37)	1.58 (2.12)	1.05 (2.04)
	Working Memory	5.12 (4.90)	3.70 (3.65)	2.00 (3.16)
	Plan/Organize	3.83 (4.31)	2.82 (3.05)	1.38 (2.96)
	Task-Monitor	3.86 (3.63)	3.58 (2.73)	2.90 (3.55)
	Organization of Materials	1.76 (2.34)	.88 (1.44)	.33 (.58)
	CRI	16.83 (16.13)	12.57 (11.17)	7.67 (11.36)
	GEC	29.15 (27.02)	20.63 (18.88)	14.90 (20.50)

Table 6

Bivariate Correlations among Three Subscales of the BRIEF2

Scale	BRI	ERI	CRI
1. BRI	—		
2. ERI	.727**	—	
3. CRI	.652**	.648**	—

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7

Bivariate Correlations among the Five EYT Tasks

Scale	1	2	3	4	5
1. GO Task	—				
2. NO-GO Task	.281**	—			
3. NOT THIS Task	.085	.287**	—		
4. MR.ANT Task	.174*	.309**	.253**	—	
5. CSORT Task	.053	.210*	.273**	.225**	—

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 8

Bivariate Correlations among the Total Score of the BRIEF2 and the EYT Tasks in Each Group

	BRIEF2 Score					
EYT Score	1	2	3	4	5	6
1. Full sample	-.380**					
2. Male		-.405**				
3. Female			-.370**			
4. Grade 1				-.346**		
5. Grade 2					-.386**	
6. Grade 3						-.024

**. Correlation is significant at the 0.01 level (2-tailed).

Table 9

Rotated Components Matrix for the EYT Tasks in Two Factors

Variable	Component 1	Component 2
"Go" Task	-.120	.872
"No-Go" Task	.407	.636
"Not This" Task	.723	.120
"Mr. Ant" Task	.516	.418
"Card Sorting" Task	.756	-.041

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization

Appendix B

Figures

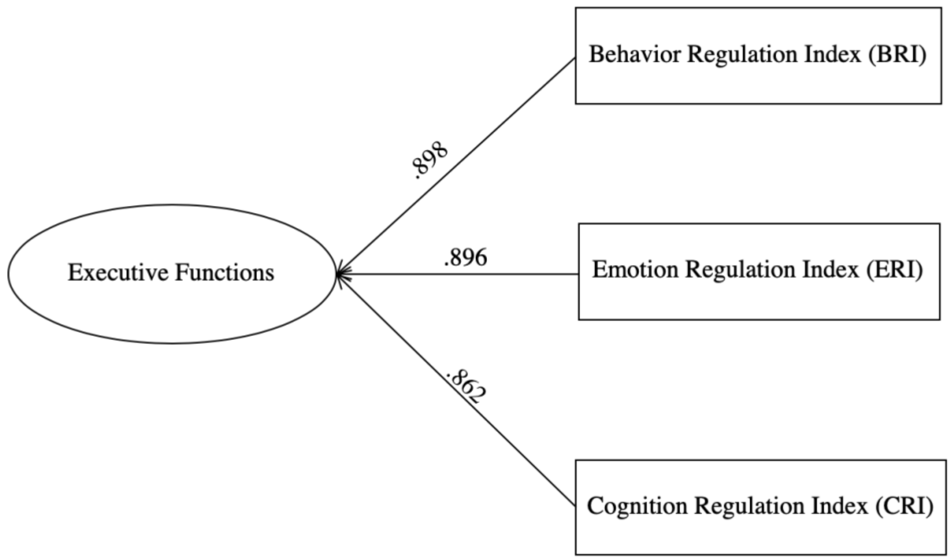

Figure 1. The Path Diagram of the BRIEF2 for the Full Sample

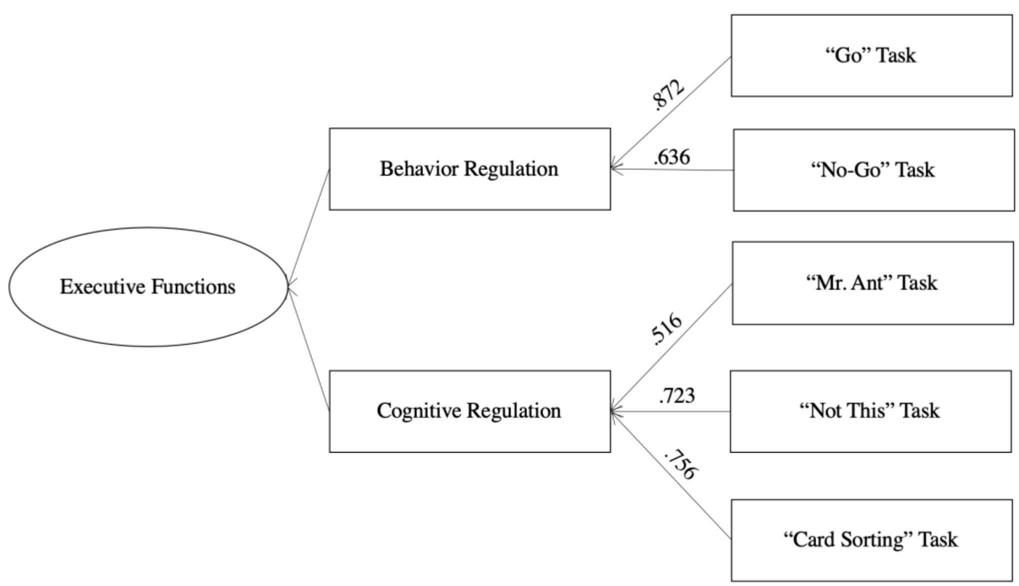

Figure 2. The Path Diagram of the EYT for the Full Sample (Subscales)

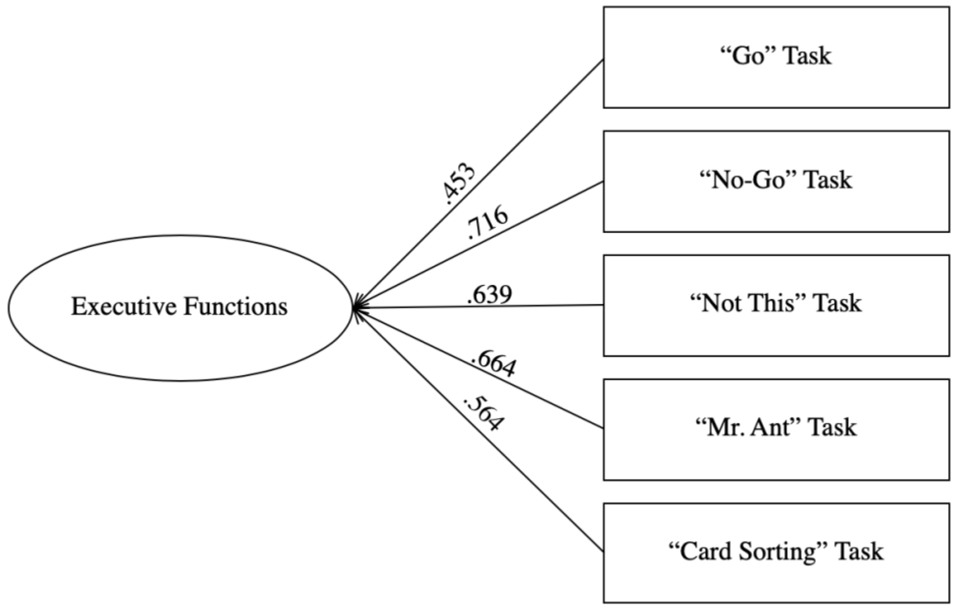

Figure 3. The Path Diagram of the EYT Tasks for the Full Sample (Global)

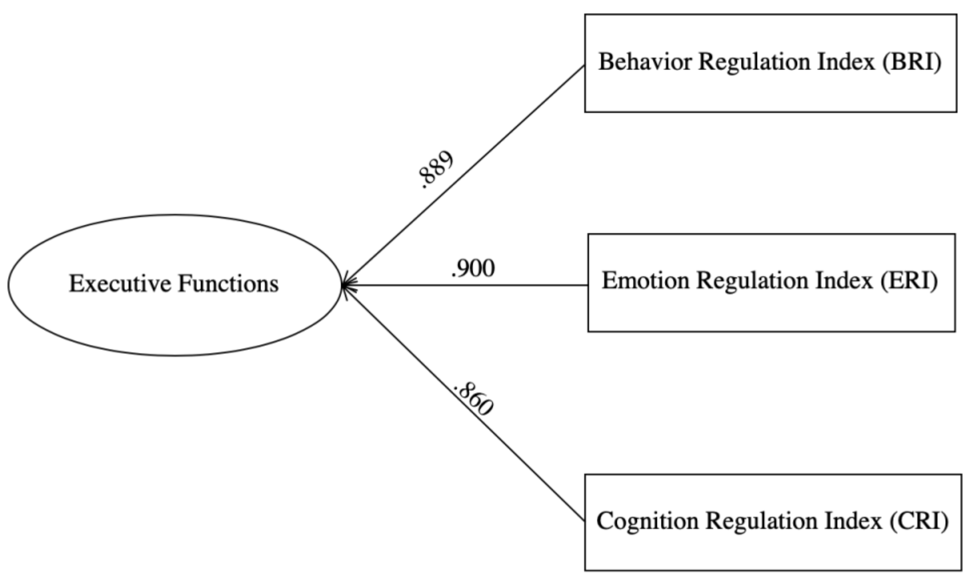

Figure 4. The Path Diagram of the BRIEF2 for the Males

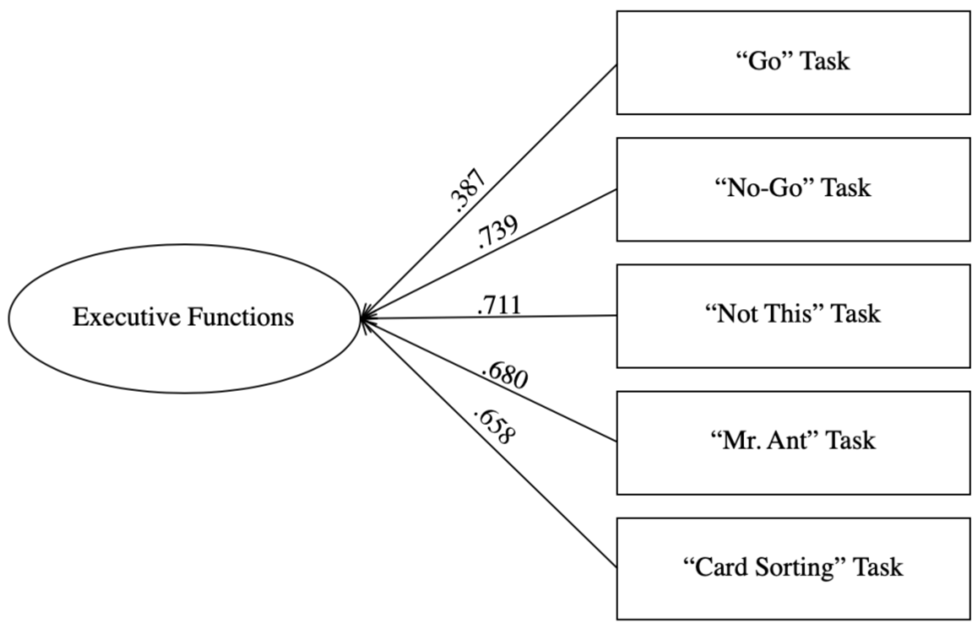

Figure 5. The Path Diagram of the EYT Tasks for the Males

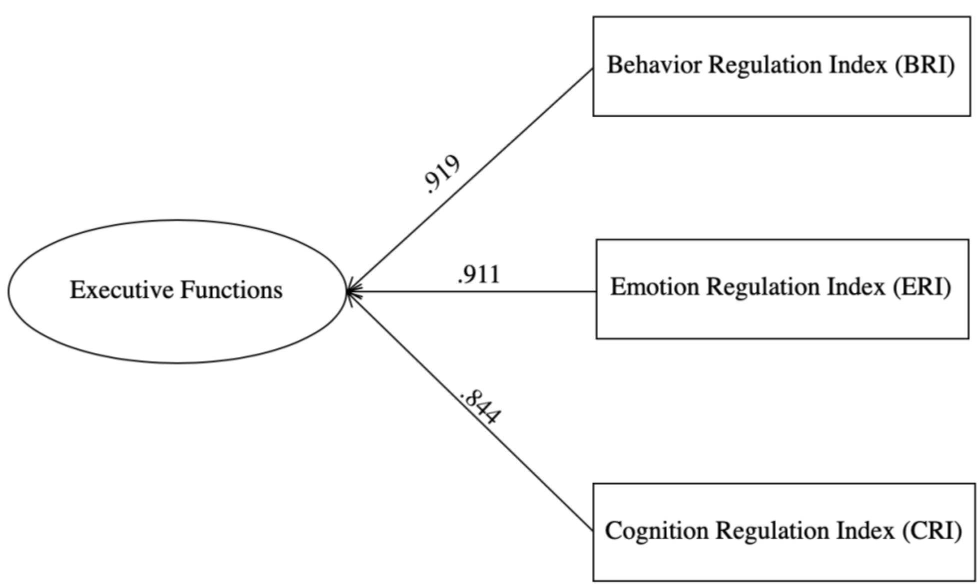

Figure 6. The Path Diagram of the BRIEF2 for the Females

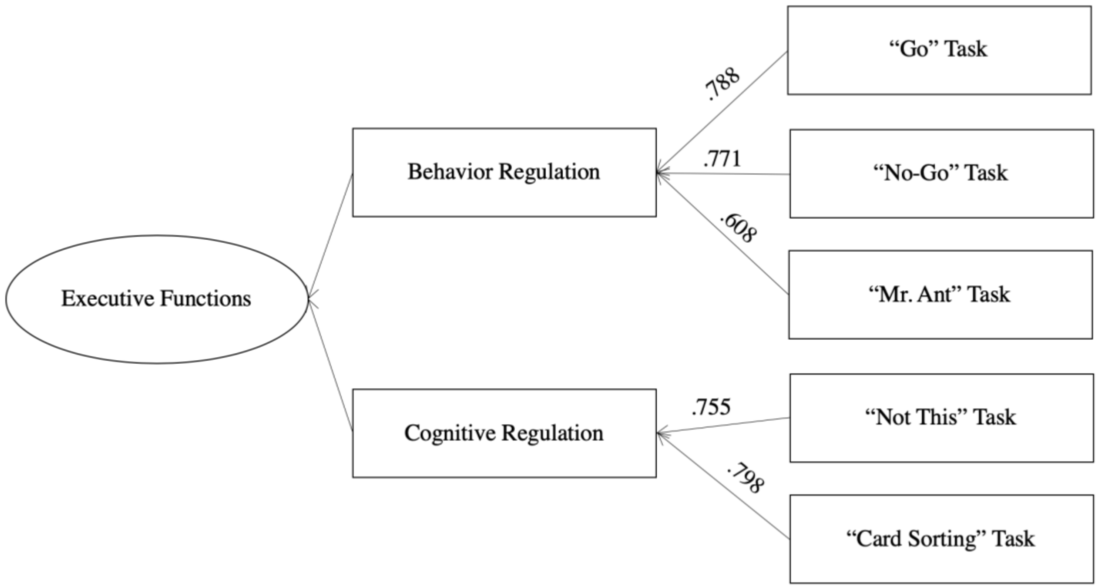

Figure 7. The Path Diagram of the EYT Tasks for the Females (Subscales)

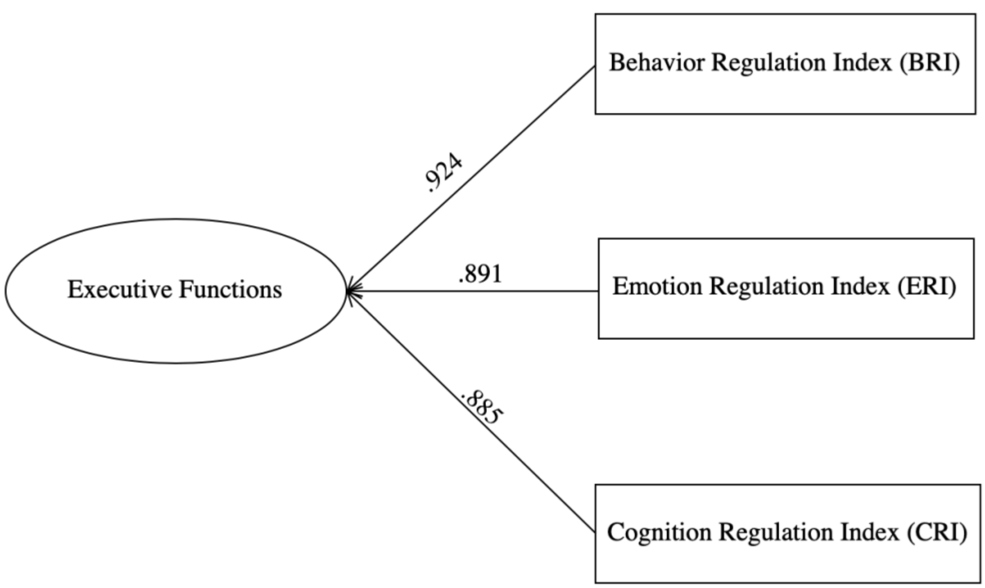

Figure 8. The Path Diagram of the BRIEF2 for the First-Grade Students

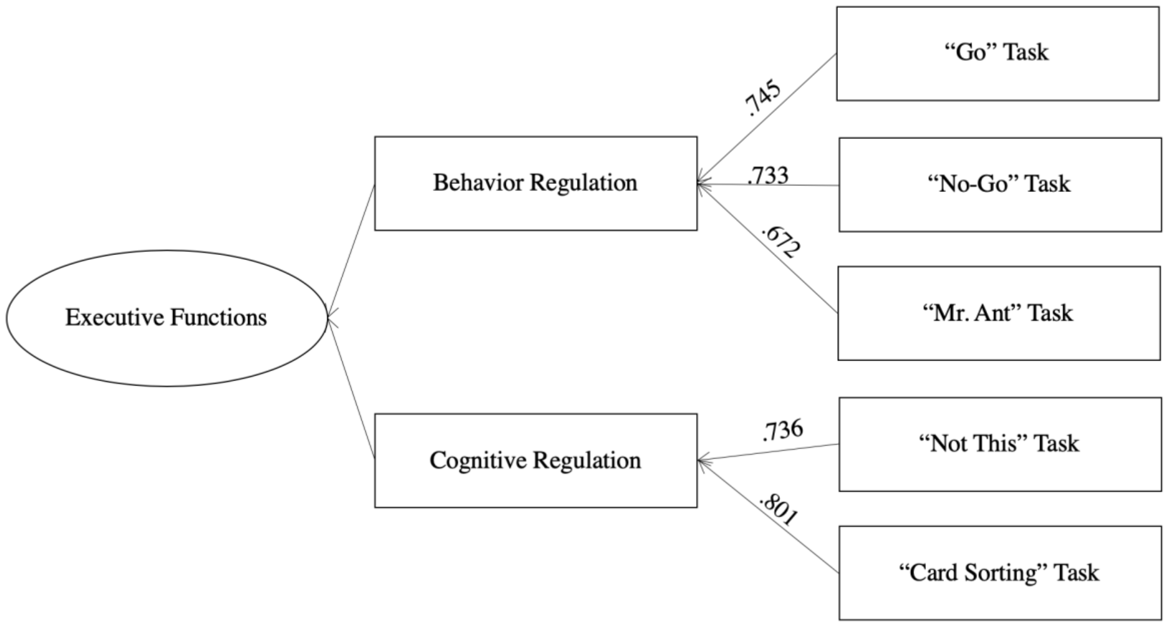

Figure 9. The Path Diagram of the EYT Tasks for the First-Grade Students

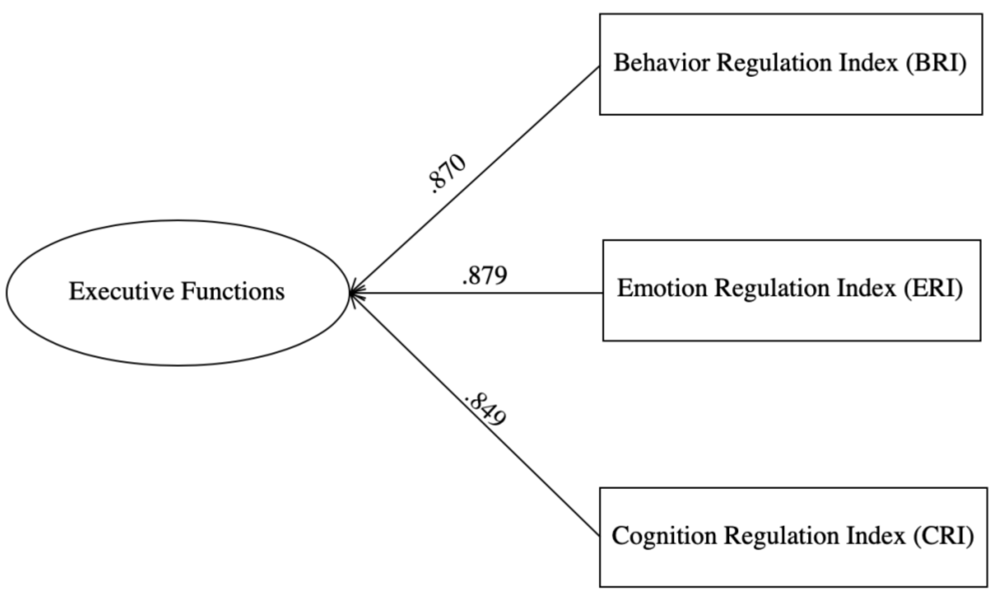

Figure 10. The Path Diagram of the BRIEF2 for the Second-Grade Students

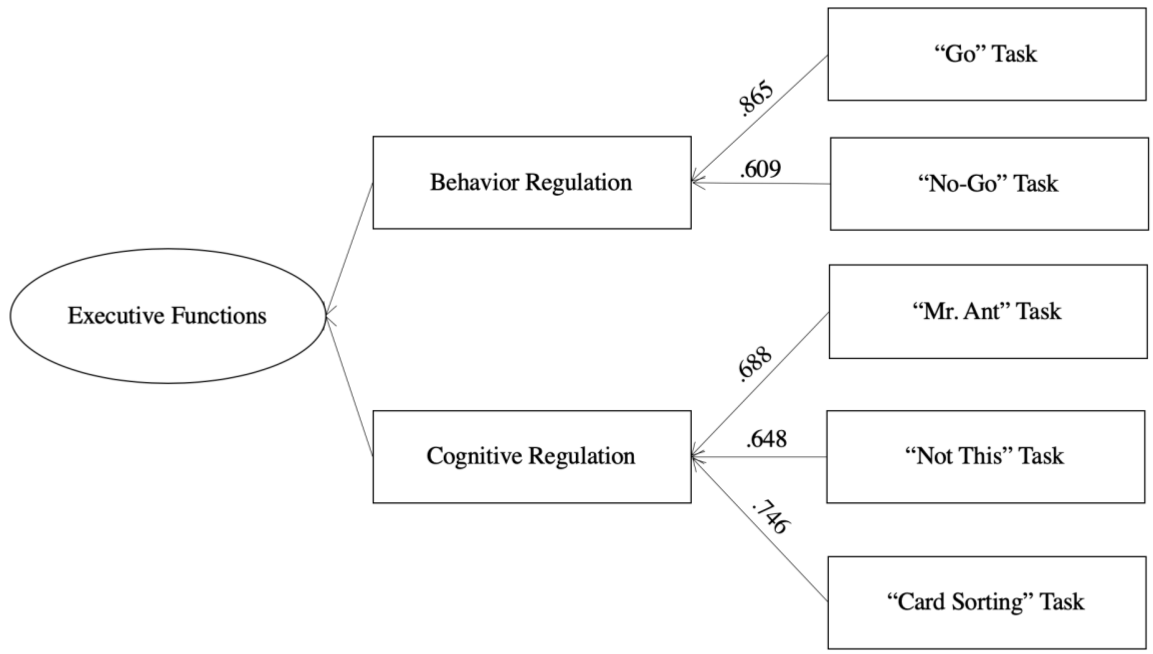

Figure 11. The Path Diagram of the EYT Tasks for the Second-Grade Students (Subscales)

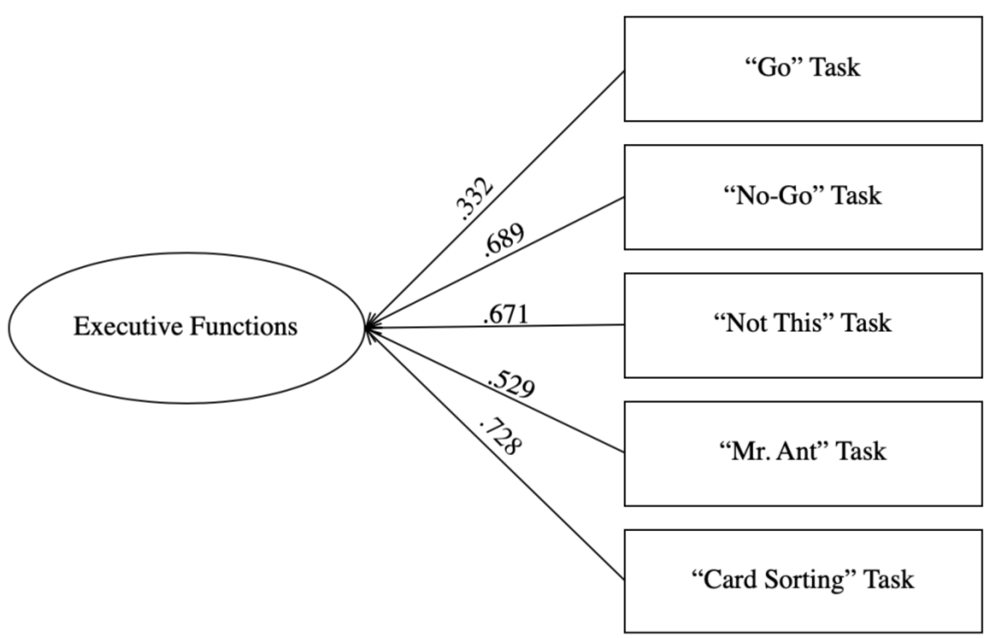

Figure 12. The Path Diagram of the EYT Tasks for the Second-Grade Students (Global)

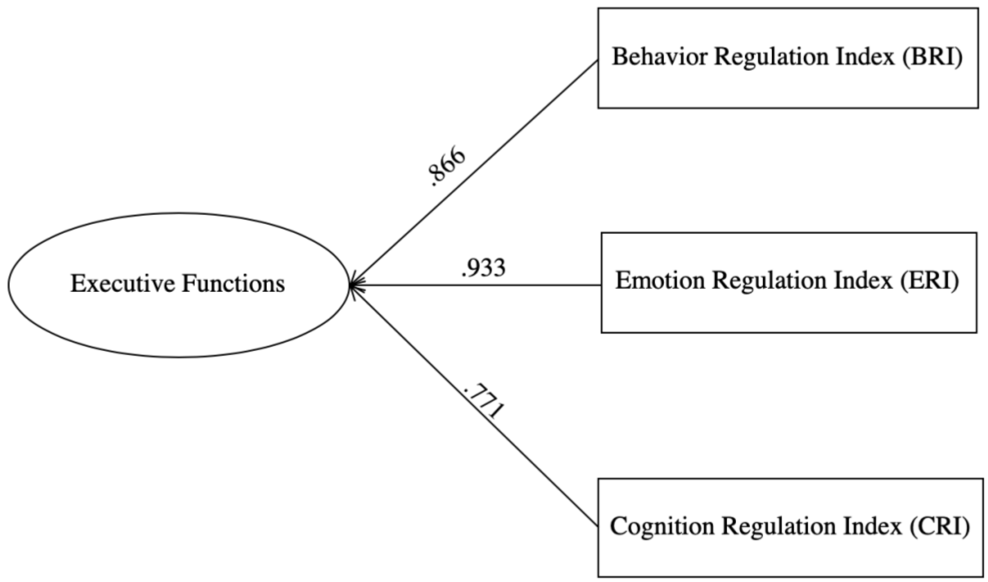

Figure 13. The Path Diagram of the BRIEF2 for the Third-Grade Students