

The Relation Between Preschoolers' Executive Functioning and Their Everyday Behaviors

by

Dana Paula Liebermann
Hon.B.Sc., University of Toronto, 2002

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

© Dana Paula Liebermann, 2006
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

The Relation Between Preschoolers' Executive Functioning and Their Everyday Behaviors

by

Dana Paula Liebermann
Hon.B.Sc., University of Toronto, 2002

Supervisory Committee

Dr. Ulrich Mueller, (Department of Psychology)

Supervisor

Dr. Kimberly A. Kerns, (Department of Psychology)

Departmental Member

Dr. Christopher E. Lalonde, (Department of Psychology)

Departmental Member

Dr. Joan Morgan Martin, (Department of Educational Psychology & Leadership Studies)

Outside Member

Dr. Michael E. Masson, (Department of Psychology)

Additional Member

Supervisory Committee

Dr. Ulrich Mueller

Supervisor

Dr. Kimberly A. Kerns

Co-Supervisor or Departmental Member

Dr. Christopher E. Lalonde

Departmental Member

Dr. Joan Morgan Martin

Outside Member

Dr. Michael E. Masson

Additional Member

Abstract

Executive Functioning (EF), a critical component of children's cognitive development, is often assessed via measures adapted from neuropsychology which impose limitations on relating EF abilities to children's everyday behaviors. A battery of EF task that emphasized specific components of 5 subscales of a behavior rating scale of EF (Behavior Rating Inventory of Executive Function – Preschool Version; Gioia, Espy, & Isquith, 2004) was administered to 60 preschool-aged children while data from both parents and teachers were collected for the BRIEF-P to clarify the meaning of traditional empirical measures of EF and the value of assessing the EF abilities of children via their everyday behaviors. Results for the EF tasks are consistent with previous findings of age related changes and relations among EF tasks. However, although the construct validity of the BRIEF-P was validated, a lack of relations between EF tasks and BRIEF-P ratings suggests that these two methods may be assessing EF abilities in distinct manners that cannot be compared.

Table of Contents

Abstract.....	iii
Table of Contents.....	iv
List of Tables.....	v
Acknowledgments.....	vi
Introduction.....	1
Method.....	14
Results.....	23
Discussion.....	31
Developmental Trends of Measures of EF.....	31
Relations Between EF Measures.....	34
Construct Validity of the BRIEF.....	35
Relationships between the BREIF-P and measures of EF.....	37
Conclusion.....	41
References.....	43
Appendix A: Tables.....	50

List of Tables

<i>Table 1:</i> Descriptive Statistics & Performance on All EF Measures as a Function of Age.....	47
<i>Table 2:</i> Means of Emotion Ratings for Children's Emotional Regulation.....	48
<i>Tables 3a & b:</i> Correlations between EF measures.....	49
<i>Table 4:</i> Internal Consistency (α) for Parent and Teacher Ratings on BRIEF-P.....	50
<i>Table 5:</i> Correlations Between BRIEF-P Parent and Teacher Ratings.....	50
<i>Table 6:</i> Factor Loadings Three-Factor Model for Parent and Teacher Ratings on BRIEF-P.....	51
<i>Table 7:</i> Correlations between EF measures and BRIEF-P Parent and Teacher Ratings.....	52
<i>Table 8:</i> Partial Correlations between EF Constructs and BRIEF-P Parent and Teacher Ratings.....	53
<i>Table 9:</i> Correlations between EF measures and BRIEF-P Ratings.....	54
<i>Table 10:</i> Correlations between EF measures and BRIEF-P Parent and Teacher Factors.....	55

Acknowledgements

With gratitude, I thank:

Dr. Ulrich Müller, my mentor and supervisor, for his insights, encouragement and continued guidance.

Dr. Kimberly Kerns, Christopher Lalonde, and Michael Masson, for their helpful suggestions, and Dr. Joan Martin for serving as the external member of my committee.

Dr. Philip D. Zelazo, for providing the foundation of my career in developmental psychology.

My mother, father, and grandmother, for their support and whose examples serve as continuing inspirations.

Marty Finestone, for having confidence and completing the picture.

The Relation Between Preschoolers' Executive Functioning and Their Everyday Behaviors

Executive Function (EF) is an umbrella term that generally refers to the mental operations involved in the conscious control of thoughts and actions (Baddeley, 1996; Perner & Lang, 1999; Stuss & Knight, 2002). For example, EF is often used to describe various cognitive processes that are involved in disengagement from prepotent stimuli and engagement in weaker stimuli. Traditionally, impairments in EF have been linked to lesions of the prefrontal cortex in adults (Stuss & Knight, 2002). Over the past two decades, a considerable amount of research has moved beyond the assessment of EF in adults and has studied the development of EF in both typically and atypically developing children. Most research studies have examined the development of EF in children by adapting measures from cognitive neuropsychology. These measures have been characterized as simply “downward extensions of measures that were originally developed for use with adults” (Donders, 2002, p. 229). This way of assessing EF dominates the literature in developmental psychology as there has, until recently, been no other means of measuring EF in children. As a result, there are several issues that remain unclear with regards to EF development in children, including the relation between children’s performance on EF tasks and their everyday behavior.

Recently, Gioia, Espy, and Isquith (2003) have modified the original Behavior Rating Inventory of Executive Functioning (BRIEF; Gioia, Isquith, Guy, & Kenworthy 2000) for use with 2- to 5-year-olds as a means to assess executive functions in everyday contexts. This new rating scale of children’s behavior in the natural environment, the BRIEF Preschool version (BRIEF-P), may provide useful information about EF in children’s everyday behavior, but it is unclear how and whether ratings on the BRIEF-P are related to performance on EF tasks. The major goal of the present study is to examine the relation

between preschoolers' performance on EF tasks and their BRIEF-P rating. The results from the present study will contribute to clarifying the ecological validity of neuropsychological EF assessments as well the meaning of the BRIEF-P.

In the following sections, executive functioning is first discussed from the neuropsychological perspective. Within this perspective, research has identified different components or aspects of executive functioning. Although findings are not entirely consistent (Zelazo & Müller, 2002), recent studies converge on identifying three major aspects of EF: Shifting, working memory, and inhibition. Each of these different aspects is briefly discussed. Next, the Behavior Rating Scale of Executive Functioning Preschool version is introduced. The BRIEF-P is comprised of 5 scales: Shifting, working memory, inhibition, planning, and emotion regulation. Research regarding the latter two scales is summarized. Finally, the hypotheses of this study are presented.

Executive Functioning

Research on the development of EF has established certain key facts: (a) EF most likely emerges around the end of the first year (b) EF continues to develop across a wide range of ages with adult-level performance being reached by about 12 years of age on most tasks, and (c) failures of EF occur in different situations at different ages (Zelazo & Müller, 2002). Despite the numerous efforts to define the precise nature of EF, the meaning of EF still remains elusive (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003). Some researchers have conceptualized EF to be a single entity that could be comparable, on a conceptual level, to the general intelligence factor of "g" (Lehto et al., 2003; Duncan, Burgess, & Emslie, 1995). Other researchers (e.g., Baddeley, 1996, Robbins, 1998) have preferred to take a multi-componential or fractionated view of EF, believing that EF includes a number of independent subfunctions. The fractionation of EF has been investigated by administering

and then factor-analyzing comprehensive neuropsychological batteries (Zelazo & Müller, 2002). Results suggest that there are dissociable dimensions of EF, a viewpoint in line with early lesion studies – a previous effort to “fractionate” EF (e.g., Robbins, 1998; Zelazo et al., 2003). Conceptualizing EF as a set of basic functions has the advantage that these functions can be operationally defined in a precise manner.

There are several reasons why it is important to gain a clear understanding of the subfunctions that comprise EF and the relations between them. Interpreting the results from EF tasks is dependent on the specifications of the task demands, which can only be decided upon once the various EF components have been determined. In addition, in order to adapt adult EF tasks into age-appropriate versions for children it is necessary to develop them along the same components. Children can be considered the ideal candidates to measure the relations between various components of EF due to their relatively limited executive capacities (Hughes & Graham, 2002). Experimental manipulations may have a stronger effect on children’s performance because their cognitive systems are easily over-loaded, and therefore “interactive effects of reciprocal causation can be investigated directly using within task designs” (Hughes & Graham, 2002, p. 139).

Studies that have examined the componential structure of EF by administering a battery of EF task and subjecting the data to factor analysis have not always yielded consistent results (see Zelazo & Müller, 2002). In a recent study with adults, Miyake et al. (2000) suggested that EF consists of at least three basic functions: shifting, updating, and inhibition. Indeed, several developmental studies have identified these same functions as being key components in the development of executive functioning (e.g., Hughes, 1998; Lehto et al., 2003; Welsh et al., 1991). As these functions are considered to be core components of EF, they are discussed next.

Shifting. Shifting, or “attention switching”, refers to the ability to change back and forth between multiple tasks, mental sets, and operations (Miyake et al., 2000; Monsell, 1996). Since Gestalt psychologists made problem solving a central topic of research in the early part of the 20th century, inflexibility has been regarded as a major hurdle to overcome when trying to solve a problem (Zelazo et al., 2003). A common explanation of inflexibility is that it involves the failure to disengage from salient or prepotent stimuli (Kirkham, Cruess, & Diamond, 2003). For example, in the Dimensional Change Card Sort (DCCS; Frye, Zelazo, & Palfai, 1995), a commonly used measure of EF in preschoolers, children are unable to inhibit a tendency to sort cards by a particular set of rules. In the standard version of the DCCS, children are presented with two target cards (i.e., blue rabbit and red boat) that vary on two dimensions (e.g., color and shape), then they are shown test cards (e.g., a red rabbit and a blue boat) can be sorted in two ways (i.e., by shape and color). Children are asked to sort by one rule, then switch to the other rule. Until the age of 4, children typically continue to sort by the first dimension, regardless of which dimension is presented first and despite being told the new rule on every trial and having correctly sorted by the “new” rule on previous occasions (Zelazo et al., 2003).

More recent work suggests that this conceptualization of inflexibility may in fact be too simple and that shifting also requires the ability to overcome proactive interference (i.e., negative priming) to perform a new operation (Zelazo et al., 2003). Furthermore, developmental research has shown that “inflexibility occurs in different contexts at different ages” (Zelazo et al, 2003, p. 4). For example, 2-year-old children persevere on single rule tasks, while 3-year-olds are able to integrate 2-rules but have difficulty with higher order rule structures (Zelazo & Müller, 2002). In addition, among 3- to 4-year old children representational inflexibility is manifested in their false belief understanding (e.g., Carlson &

Moses, 2001), reasoning about causality (e.g., das Gupta & Bryant, 1989), and reasoning about delayed representations (e.g., Zelazo, Sommerville, & Nichols, 1999).

Updating or Working Memory. The second target EF function that Miyake et al. (2000) identified, updating, refers to the updating and monitoring of representations held in working memory (WM). This function requires the monitoring and coding of incoming information that is relevant to the task an individual is faced with, followed by the revision of items held in WM and a replacement of old irrelevant information with new relevant information (Miyake et al., 2000; Morris & Jones, 1990). The term working memory is not used consistently by different authors, but is most often defined as the simultaneous manipulation and maintenance of a representation so that the representation can guide an individual's behavior (Zelazo et al., 2003).

Daneman and Carpenter (1980) used the term WM to refer to a system with limited capacity, or activation, that must deal with the burdens imposed by processing and storing information. Thus, if the amount of activation that an individual has available is not enough to meet the demands of a task, there will be a trade off between processing and storage. Studies with adults have shown that there are individual differences in activation levels which can be evaluated through the use of complex working memory tasks (Gathercole, 1998). Examples of such tasks include reading or listening span tasks where participants need to process incoming information while remembering previously encoded information, thus requiring both the storage and transformation of material. Children's performance on these complex WM tasks has been found to improve with age (Gathercole, 1998; Zelazo et al., 2003). Overall, a key aspect of WM tasks is that they require the active maintenance of information in the midst of concurrent processing and interference, therefore requiring an EF mechanism that combats the interference (Conway, Kane, & Engle, 2003). Thus, WM is

often thought to be closely linked with the concept of inhibitory control (Bull & Scerif, 2001). The relationship between WM, inhibition and their relation to EF will be further discussed below.

Inhibition. Inhibition, identified by Miyake et al. (2000) as the third EF target function, is the ability to suppress dominant, automatic, or prepotent responses. A prototypical example of an inhibition task is the Stroop task in which individuals need to overcome the tendency to produce the more automatic responses of naming a color word as opposed to the color of the ink in which the word is written. The Day/Night task is a child-appropriate version of the Stroop-task (Gerstadt, Hong, & Diamond, 1994). This task involves the inhibition of automatic responses because it requires children to say the opposite of what they see on a picture card. For example, when they see a picture of a sun they are instructed to say “night” and when they see a picture of a moon they must say “day”. This task is challenging to 3- to 4-year-olds, but children between 6- to 7-years complete the task without difficulty (Gerstadt et al., 1994). Younger children’s failure on the Day-Night task is attributed to their inability to inhibit saying a word that is semantically related to the picture they are seeing (Diamond, Kirkham, & Amso, 2002).

Inhibition itself is a complex construct, and several different types of inhibition have been distinguished (Nigg, 2001). In cognitive psychology, inhibition tasks are commonly divided into those that require effortful versus those which require automatic suppression of a particular response (e.g., Schneider & Shiffrin, 1977; Nigg, 2001). Research that involves examining inhibition in terms of clinical disorders, such as ADHD, requires a different distinction between types of inhibition that include motivation as a factor. Therefore, Nigg (2001) developed a 2-process conceptual framework of inhibition that includes executive and motivational inhibition: executive inhibition involves the deliberate suppression of motor

and cognitive to achieve a later, internally represented goal, and motivational inhibition is the suppression of a behavior or response that was primarily driven by fear or anxiety. Another distinction that is relevant for the present study is between tasks that involve simple delay inhibition and tasks that require both working memory and inhibition. Previous research has suggested that a distinction can be made between these types of tasks demands (Carlson, 2005). For example, tasks which are conceptualized as assess the former (i.e., primarily inhibitory demands) are those which require children to simply inhibit a prepotent response for a specific period of time. For example, in a Gift Delay task, children are asked to refrain from opening a present until they are told to do so. Tasks which are thought to involve both WM and inhibitory demands require children to inhibit prepotent responses while keeping specific rules in mind. The Whisper task requires children to say the names of cartoon characters they are shown and must remember to whisper the names instead of shouting or using a normal speaking voice. The distinction and relationship between inhibition and WM is discussed further in the following section.

Inhibition and Working Memory. A number of studies have explored the relations among components of EF “microlevel” variables (e.g., shifting, working memory, and inhibition; see Senn, Espy, & Kaufmann, 2004). The findings of these studies led some researchers to propose a link between working memory and inhibition, and this link is often examined in the context of research on the developmental relation between EF and Theory of Mind (ToM). For example, Carlson, Moses, & Breton (2002) have suggested that both working memory and inhibition contribute to mental state attribution. Carlson et al. (2002) found that a task which required children to inhibit an inappropriate prepotent response while activating a conflicting novel response related more strongly to ToM than did a task that required children simply to inhibit responding. Only the former type of task appeared to

impose substantial working memory demands, while both required the use of inhibition. The authors related the initial task to ToM tasks which “require children to suppress their knowledge of reality while activating a seemingly incompatible representation of that reality” suggesting the combination of working memory and inhibition may be vital. This follows the proposals of others who have suggested that prefrontal functions involve an interaction between inhibition and working memory (e.g., Bjorklund and Harnishfeger, 1990; Dempster, 1991; Diamond, 1991; Diamond and Taylor, 1996; Roberts and Pennington, 1996).

It has also been found that WM, inhibition and shifting are also partially related to each other in children (e.g., Senn et al., 2004), as well as in adult populations (e.g., Burgess, 1997, Lehto et al., 2003) and adolescents (e.g., Lehto, 1996). For example, Senn et al. (2004) found that 2- to 6-year-olds’ performance on a working memory task, on an inhibition task, and on a measure of mental flexibility predicted performance on a complex problem-solving task (the Tower of Hanoi).

BRIEF- P

Only a few studies have used observations from a naturalistic context to examine the contribution of EF to children’s social competence (e.g., Peskin & Ardino, 2003) and problem behaviours (e.g., Fahie & Symons, 2003). However, there has been little research relating EF to children’s everyday behaviors using a behavior rating scale of EF. Assessing everyday behaviors of children both in the home and school environment provide excellent opportunities for examining routine manifestations of EFs. The Behavior Rating Inventory of Executive Functioning (BRIEF; Gioia et al., 2000) was developed for this purpose: to further understand children’s and adolescents’ (i.e., 5- to 18-year-olds) executive function competence in a real-world setting (Baron, 2000). The BRIEF samples children’s everyday executive skills in natural settings by having parents and/or teachers complete a rating scale

of 86 items involving statements about the child's daily life, requiring a forced-choice response (i.e., "never", "sometimes", or "often"). The BRIEF "adds a complementary ecological validity dimension" to neuropsychological assessment batteries of EF (Isquith et al., 2004). Studies conducted with the BRIEF have demonstrated that it "captures profiles" of EF that differ across various disorders including Attention-Deficit Hyperactivity Disorder and Autism Spectrum Disorder (Gioia, Isquith, Kenworthy, & Barton, 2002). In 2003, a preschool version of the BRIEF (BRIEF-P) was developed. The new behavior rating scale consists of 63-items that make up 5 scales: Inhibit Shift, Emotional Control, Working Memory, and Plan/Organize. The initial results of the use of the BRIEF-P have produced a 3-factor model solution based on both parent and teacher ratings. The first factor, *Emergent Metacognition*, consists of WM, Plan/Organize, and Inhibition and reflects developing metacognition. The second, *Flexibility*, is comprised of Shift and Emotional Control, reflects behavioral rigidity and emotional modulation. Finally, items of the Inhibition scale and the Emotional Control scale loaded on the third factor, *Inhibitory Self-Control*.

In addition to Shifting, Working Memory and Inhibition, the BRIEF-P also measures two other processes: Planning and emotional control. Planning is often considered to be an aspect of working memory (Miyake et al., 2000), although some factor analytic studies have also identified a specific planning factor (Levin et al., 1991). Emotional Control itself is not necessarily considered to be a component of EF, but it is believed that emotional control is influenced and, in turn, influences the development of EF. Planning and emotion control are discussed in the next two sections.

Planning/Organization. Planning is a purposeful behavior that requires setting a goal, deciding on a plan, creating a plan, implementing and monitoring the plan, and finally reviewing the outcome (Scholnick, Friedman, & Wallner-Allen, 1997). The presence of

strategic behaviors is often used as an illustration of planning in EF in both adults and children. Tasks used to evaluate planning skills often supply elements and “moves” and the individual must determine the correct sequence in accordance with certain explicit and implicit restrictions (Scholnick et al., 1997). Although young children do not demonstrate overt behaviors that are considered “planful” in nature, preschoolers do possess knowledge of the sequence of routine events. From the age of 3- to 5-years, there is considerable development with regards to children’s abilities to reflect on sequences of events and focus on planning-related components (Hudson, Sosa, & Shapiro, 1997). Previous research has also shown that preschoolers’ ability to construct plans in advance of performing activities develops before their abilities to execute such plans (Hudson et al., 1997). The basis of this discrepancy is attributed to the absence of requisite cognitive skills for remembering and monitoring plans while they are being carried out. Such cognitive skills are gained as various skills related to EF develop: more complex knowledge of sequences of events, more flexibility in task performance allowing for reflection, and experience in plan construction and execution (Hudson et al., 1997).

A commonly used measure of planning when studying children is the Tower of Hanoi (TOH), which requires participants to transfer disks on pegs from an initial start state to a goal end-state in a limited number of moves and by abiding to specific rules (Bull, Espy, & Senn, 2004). Although other cognitive processes (e.g., WM and shifting) are thought to contribute to successful performance on the TOH, it is generally described as a higher-order planning task. Extensive research using the TOH has demonstrated significant developmental shifts in performance from ages 3- to 4-years to 5- to 6-years, with adult level skill being obtained in complex move problems by adolescence (Welsh et al., 1991).

Emotional Control/Regulation. Recently, the concept of emotion regulation has received considerable attention (e.g., Bradley, 2000; special issue of *Child Development*, March/April 2004). At the same time, there is some concern in the field that the concept has become too wide and has lost any specific meaning. The study of emotion regulation is complicated by the fact that it is often difficult to distinguish the intensity of an initial emotional reaction from the regulation of that emotion (Kagan, 1994; Cole et al., 2004). Traditional emotional measures, such as the study of facial expression or physiological markers, are thought to reflect both regulatory influences as well as the emotions themselves. There is little doubt, however, that emotional regulation is an important concept, and that competent emotion regulation strongly influences healthy psychological functioning.

Previous research on emotional regulation in young children can be classified into three basic categories based on their different foci: infant temperament, mother and child face-to-face interaction, and general early childhood emotional self-regulation. Based on the various methods used to test the above categories (e.g., fear-activation procedures, reactions to facial and vocal cues) a developmental sequence for emotional regulation has been mapped. During infancy there is a basic self-regulatory capacity for managing various emotions as 6-, 12-, and 18-month-olds have been shown to engage in various degrees of regulatory behavior depending on their temperament (Mangelsdorf, Shapiro, & Marzolf, 1995). This is followed by a period where infants are able to engage in mutually regulatory interactions with others, which can be accomplished through vocalization, gesturing and communicating distress (Cole et al., 2004). Additional self-regulating strategies, such as refraining from displaying negative emotions when appropriate, are developed over the toddler and into the preschool years (Cole et al., 2004). The development of executive functioning is believed to be an important contributor to this last stage, with two types of

cognitive skills contributing to the development of the self-control of emotion. First, preschool children develop cognitive self-regulation or the knowledge certain behaviors are more useful than others in particular situations. This type of regulation involves planfulness, control, reflection, competence, and independence (Paris & Newman, 1990). Second, through increased social interaction, children acquire the understanding that individuals they interact with will respond or behave in a particular manner. From this they will gain insight as to when it is necessary and appropriate to regulate displays of affect (Fox & Calkins, 2003).

Recent research in neuropsychology has suggested that EF abilities may operate differently under contexts that are more affective in nature (e.g., Bechara, 2004; Hongwanishkul et al., 2005). Zelazo & Müller (2002), introduced the term “hot” EF to refer to aspects of EF that are elicited by problems that require the regulation of affect or motivation, while “cool” EF is required for more abstract, decontextualized problems (Hongwanishkul et al., 2005). Although it is probable that most EF tasks require a combination of both hot and cool as they are both thought to be a part of a single coordinated system, this distinction helps to highlight the role influence emotions and their regulation may have on the performance and hence the evaluation of EF abilities.

Goals of the Study

The majority of previous studies that utilized the BRIEF have examined the correlations between the BRIEF and laboratory assessments of EF in clinical populations (e.g., Anderson, Anderson, Northam, Jacobs, & Mikiewicz, 2002). The arrival of the BRIEF-P permits a similar comparison of the relations between typical preschoolers’ performance on EF tasks and their behavior ratings of EF. The primary purpose of this study is to examine relations between EF tasks (e.g., Tower of Hanoi) that emphasize

specific components of EF and the 5 subscales of the BRIEF-P (Working Memory, Planning/Organization, Inhibition, Shifting, and Emotional Control) in 3-, 4-, and 5-year-olds.

The present study will explore the relations between children's performance on EF tasks and their behavior as rated on the behavior rating scale. In this manner, the study will shed light on the ecological validity of EF tasks, and will clarify the meaning of the different scales of the BRIEF-P. The results will be interpreted in terms of predictions that the subscales of the BRIEF-P correlate significantly with EF tasks that differentially emphasize these components. In addition, developmental trends within each EF task and the BRIEF-P will be examined. Specifically, the following relations are expected to be significantly correlated: the Disappointment Procedure is expected to correlate with the Emotional Control scale, the Whisper and Gift Delay will be related with the Inhibition scale, the ID/ED task will be linked to the Shift scale, both the Forward and Backward Digit Span tasks will correlate with the WM scale, and the Tower of Hanoi will be related to the Plan/Organize scale. Other, weaker, correlations are hypothesized to be found between the Disappointment Procedure and Inhibition as children will be required to inhibit negative emotions such as anger and sadness. Inhibition may also be related to some extent to the Backward Digit Span task as children will need to inhibit the automatic tendency to repeat the series of numbers in the sequence that was said by the puppet (i.e., in same order as opposed to in reverse). The Tower of Hanoi may be weakly correlated to the Shift as mental flexibility is considered to be a cognitive process that contributes to successful performance on this task (Bull et al., 2004). And finally, the performance on the ID/ED Shift task may be related to the WM scale as the task requires that children remember previous picture selections that were made in order to try establish what rule it is they should be following.

It is predicted that relations among EF tasks will be similar to those found in previous studies. Specifically, correlations are anticipated between the measures of inhibition (i.e., Gift Delay and Whisper task) and WM (i.e., Forward and Backward Digit Span). Also, measures that require some degree of flexibility (i.e., Disappointment Procedure, ID/ED Shift task, and Tower of Hanoi) are expected to correlate with measures of inhibitory control (i.e., Gift Delay and Whisper task) and with each other. As a result of these correlations it is hypothesized that the EF tasks will not produce the same 3-factor model as the BRIEF-P. It is hypothesized the factor of “Flexibility” will most likely also include aspects of inhibition as they have been found to be related frequently in the literature (e.g., Zelazo et al., 2003), as well as aspects of Planning/Organizing as tasks that involve flexible thinking require children to set a plan of action when trying to achieve a goal. (Zelazo, Carter, Reznick, & Frye, 1997).

A measure of verbal ability, the Peabody Picture Vocabulary Test (PPVT-3; Dunn & Dunn, 1997) will also be administered to participants and the results will be examined in relation to their EF and BRIEF-P scores. Previous studies have shown that receptive vocabulary scores, as measured by the PPVT-3, correlated with measures of EF (e.g., Fahie & Simons, 2003). In the present study, the relation between receptive vocabulary and performance on EF tasks as well as EF behavior ratings will also be examined, as well as whether relations between corresponding components of EF tasks and the BRIEF-P remain significant even after controlling for receptive vocabulary.

Method

Participants

Sixty preschool children between the ages of 37 and 70 months (23 boys and 37 girls, $M = 52.55$ months, $SD = 8.16$) undergoing normative development from a medium sized

city in Canada participated in the study. Children were recruited from various local preschool and child care facilities. An additional four children were dropped from the final sample (mean age = 54 months) because of refusal to complete all stages of testing ($n=2$), uncertainties regarding normative development ($n=1$), or insufficient command of the English language ($n=1$). The remaining sixty participants were divided into two age groups: Younger preschoolers (14 boys and 16 girls, $M = 45.47$, $SD = 3.82$) and older preschoolers (9 boys and 21 girls, $M = 59.63$, $SD = 4.16$).

One parent and one teacher provided ratings for each of the children. For one participant neither parent nor teacher provided a rating, and for another participant a parent did not provide a rating. The majority of parents who responded (98%) were the mother of participants and all of the teachers who responded were the children's primary caregiver in their daycare or preschool.

Procedures

All participants were tested individually by a female experimenter at their respective daycare or preschool and all sessions were videotaped. The tasks were administered during a 1 hour session in the following standardized order: Whisper, Intradimension/Extradimension shifting, Backward Digit Span, Forward Digit Span, Gift Delay, Tower of Hanoi, Peabody Picture Vocabulary Test (PPVT-3; Dunn & Dunn, 1997), and the Disappointment Procedure. The large number of empirical tasks used and the need to separate tasks of similar format prohibited administering the battery in a counterbalanced fashion. The children received three toys worth approximately \$5 for their participation and the participating daycares and preschools received a gift worth approximately \$10.

Measures

BRIEF-P Subtasks (Gioia, Espy, Isquith, 2003)

1. Emotional Control. This dimension measures a child's ability to modulate their emotional responses. Children with difficulties in this area (i.e., poor emotional control) can have overblown reactions to both positive and negative events. Examples of BRIEF-P items related to emotional control are "Becomes upset too easily" and "Mood Changes Frequently".

2. Inhibit. This dimension measures the child's ability to resist responding or acting on an impulse (i.e., inhibit) and to stop their behavior at the appropriate time. Some have argued (e.g., Burgess, 1997) that poor inhibition is the general underlying ability in executive dysfunction. Also, children who have suffered traumatic brain injuries show poor inhibitory skills and increased impulsive behavior. Items on the BRIEF-P related to inhibition include such items as "Is impulsive" and "Acts too wild or out of control".

3. Shift. This dimension measures a child's ability to move from one situation to another (e.g., from one aspect of a problem to another) as is required by the current conditions. The key aspects of this scale include abilities to make transitions, switching or alternating attention, problem solving flexibility, and changing focus from one mindset or topic to another one. Mild deficits are displayed in poor problem solving efficiency, while more severe discrepancies can be reflected in perseverative behaviors. BRIEF-P items include "Is upset by change in plans or routines", "Becomes upset with new situations", and "Has trouble with activities that involve more than one step".

4. Working Memory. This dimension measures the child's ability to hold information in their mind for the purpose of completing a particular task. WM is considered vital for carrying out multiple activities, implementing a series of actions, or following complicated instructions. BRIEF-P items related to WM include "Cannot stay on the same topic when talking" and "Has trouble remembering something, even after a brief period of

time”. WM is often associated with the abilities to sustain attention and performance.

Although these elements have been conceptualized as distinct, the BRIEF-P incorporates both cognitive action and holding information actively in mind into the WM scale. Items related to these aspects include “Needs help from adult to stay on task” and “Has short attention span”.

5. Plan/Organize. This dimension measures the child’s ability to manage current and future-oriented task demands in a particular situation. Planning refers to the ability to anticipate events that will occur, executing the instructions or goals, and developing steps in preparation of performing a task or activity. In preschoolers, planning often involves implementing a goal by selecting the most efficient manner to achieve that goal, which often involves sequencing a series of steps together. BRIEF-P items related to this aspect include “Has trouble thinking of a different way to solve a problem or complete an activity when stuck”. Organizing refers to the ability to bring together the information and actions necessary to achieve the goal at hand. The manner in which such material is organized can have a large effect on how the information is learned, remembered and implemented in the future. Children who lack organizational skills will approach a problem haphazardly and be overwhelmed when presented with large amounts of information. Items in the BRIEF-P that measure this component include “Gets caught up on the small details of a task or situation and misses the main idea”.

Peabody Picture and Vocabulary Test (PPVT-3; Dunn & Dunn, 1997)

The PPVT-3 correlates highly with full-scale verbal intelligence measures such as the Wechsler Preschool and Primary Scales of Intelligence - Revised (WPPSI-R) (Carvajal, Parks, Logan, & Page, 1992) and the verbal subscale of the Stanford–Binet IV (Hodapp, 1993). The experimenter reads aloud a word, and children select the picture that best illustrates it out of

a set of four choices. The task continues children make an error on 8 out of a set of 12 words.

Executive Function Tasks

Measure of Emotional Control:

Disappointment Procedure. The participant was asked by the experimenter to rank 8 small prizes before starting a filler task administered by the experimenter. The participant was told that they needed to place the various prizes in labeled buckets ranging from a happy faced bucket to a sad faced bucket. Among the 8 prizes were some undesirable toys, including items that are visibly broken. The participant's rank-ordered prizes were placed on a tray where the first and last choices were adjacent to each other (i.e., happy faced bucket next to sad faced bucket). The experimenter placed the tray out of the participant's view administered a filler task, and when 4 minutes elapsed she told the participant that they would be awarded a prize. The experimenter returned with the participant's first choice prize in a gift bag, placed it on the table, sat down, maintained regular eye contact and a neutral facial expression, and gathered her papers for 20 seconds. If the participant spoke, the experimenter would paraphrase the participant in a neutral tone of voice. The experimenter then administered a second filler task, and after 4 minutes she told the participant that they would be receiving another prize. The experimenter then presented the participant with a gift bag that contained their last choice toy. As with the first gift, the experimenter maintained regular eye contact, neutral facial expression, and would paraphrase the participant in a neutral tone of voice if they spoke during the 20 seconds interval. The experimenter exclaimed that a mistake had been made and that the participant had received the wrong toy. The experimenter then said she would look for another toy, turned her back to the participant, and waited for 30 seconds before presenting the participant with a gift bag

containing their second choice toy. As with the prior presentations, the experimenter did not interact with the child for 20 seconds. The experimenter interviewed the participant to determine self-reported feelings during the disappointment, awareness of whether this feeling was known to the experimenter, and if the participant could articulate a way to mask their true emotions from the experimenter. Specifically, the participant was asked the following:

- 1) How did you feel when you got the (*last choice toy*)?
- 2) Did I know how you felt?
- 3) If no, how did you keep me from knowing? If yes, how did I know?

The experimenter explained the purpose of the task to the child and apologized again for the mistake.

Facial action, gestures, and vocal quality cues were used to determine the presence of 5 basic emotions (happiness, disappointment, disgust, surprise, & anger), general emotional states called “Overall Negative Affect” and “Overall Positive Affect”, and the level of the child’s engagement in the game. Two trained coders, who were unaware of the purpose of the study, coded the participant on based on two 20 second segments: after the participant received their first choice toy and after they received their last choice toy.

Measures of Inhibition:

Whisper. The task required participants to temper a verbal response. As a warm-up the experimenter asked the participant to whisper their name. The participants were instructed that they are going to play a game in which they needed to whisper the names of cartoon characters that they would be shown on cards. If they didn’t recognize a character, the participants were told to whisper “I don’t know”. In total, 10 characters were presented

of which 6 were easily recognizable by this age group. During the game the experimenter also whispered and gave a reminder of the rules of the game after the 5th trial (i.e., half-way).

A score of 0 is given for shouting, 1 for a normal voice or mixed shouting and whispering, and 2 for whispering. Trials in which no answer was given are excluded when using the average score for analysis (Carlson et al., 2004).

Gift Delay. The experimenter told the participant that he/she would receive a gift (e.g., a small toy ball). The participant was asked to sit in a chair facing away from the experimenter and told not to look while the present was wrapped so that it could be a “big surprise”. The experimenter wrapped the present in a noisy standardized manner for 60 seconds. A score of 0 was given if the participant turned around to peek, 1 if they peeked over their shoulder, and 2 if they did not attempt to peek. The latency to peek over the shoulder and the total number of peeks was also coded.

Measure of Shifting:

Intradimension/Extradimension (ID/ED). The task involved 3 sets of stimuli consisting of picture items that differed in color and pattern and 4 phases: compound discrimination, compound reversal, intradimensional (ID) shift, and extradimensional (ED) shift. Both compound phases consisted of a maximum of 75 trials with the same set of stimuli. The ID and ED phases had the same number of maximum trials, but contained different stimuli from the other phases and from each other.

Participants were presented with 2 pictures on a page that differ in color and pattern, and were asked to select the one that is ‘correct’. Participants were informed that they had to determine which picture was correct on their own, but would be provided feedback by the experimenter. A response of “yes” would be given if the participant was correct and “no” if they had chosen the wrong picture. Participants were informed that the pictures were

different shapes and different colors and that a rule could be followed to make the right choices. They were told that “sometimes that rule will change and you have to be ready for this, but the change won’t happen very often”.

The order of the 4 phases always remained constant: compound discrimination, compound reversal, ID, and ED. In the compound discrimination phase 2 pictures were presented that differed in color and pattern, and positive feedback was given for either selecting a particular color or pattern. In the compound reversal phase the feedback provided was reversed so that the level of discrimination that was correct in the discrimination phase was now incorrect, and the level that was incorrect was now correct. In the ID phase, a new exemplar for each dimension was introduced, and participants were required to continue to responding to the previously relevant dimension. The ED phase involved the introduction to another set of exemplars and participants were required to shift their responses to the previously (i.e., in the ID phase) irrelevant dimension for them to perform correctly. A switch to the new phase was never announced by the experimenter. For each of the 4 phases, participants who made 6 consecutive correct responses (out of the maximum of 75) were considered to have passed. The scoring included the number of trials-to-criterion and errors-to-criterion for each of the 4 phases.

Measures of Working Memory:

Backward Digit Span. Participants were asked to repeat a series of single-digit numbers backwards. The experimenter used a puppet to demonstrate how to say digits backwards. The experimenter said “We are going to play a silly game with Molly, so whatever she says, I will say it backwards. Let me show you how to play. If Molly says the numbers ‘1,2’ then I will say ‘2,1’”. Participants were asked to do what the experimenter had done. A 2-digit practice trial was administered, in which participants were corrected if they

were wrong, and the example provided was repeated. If the participant did not answer correctly after two repetitions of the rules, the game ended. The series size of the test trials increased with every successful trial (from 2 to 4 digits), with the highest level achieved being recorded as the participant's score (1-4). A score of 0 will be given to participants who fail the 2-digit series.

Forward Digit Span. The experimenter read aloud a series of single-digit number sequences that ranged from 2 to 9 digits in length. The participant was asked to repeat the series in the same order. For each sequence length there were 2 trials, and the task ended when the participants made an error on both trials. Participants' score was the number of trials that were successfully completed.

Measure of Planning/Organizing:

Tower of Hanoi. Participants were presented with 2 models that had 3 equally sized large pegs. One model was described as the experimenter's and had 3 different colored rings that were stacked on the right peg. The peg diameter on both models were graduated so that the three rings fit on one peg only if stacked from largest to smallest (i.e., bottom to top). The second model was identified as the participant's, contained the same colored and sized rings as the experimenter's, but the rings were arranged in another configuration across all 3 pegs. The participant was instructed that they must move the 3 rings, only one at a time, so that their model was in the same configuration as the experimenters. An instructional story was used to describe the goals and rules of the task (Welsh et al., 1991): "monkeys" (the rings) of different size (daddy, mommy, and baby) jumped between the "trees" (pegs), and the goal was to bring the monkeys home to sleep on their tree (i.e., achieve configuration of experimenters' model). The participant was informed of 3 rules for the task: 1) only 1 monkey could move at a time, 2) a bigger monkey could not sit on a smaller monkey, and 3)

the monkeys had to stay on the trees and not in the participants' hands. Following Bull et al. (2004), each of the six types of problems (from 2 through 7 moves) were presented for a single trial. There was one problem with 0 counter-intuitive moves, two with 1 counter-intuitive move, and three with 2 counter-intuitive moves. Trials were discontinued when a solution was reached or when 20 moves had been made. Testing stopped when participants performed 2 consecutive failures (i.e., refused to make moves or failed to make permissible moves).

Each of the 6 trials was assigned a point value based on the minimal moves required for each solution (e.g., 3 move problem = 3 points). The maximum score possible was 27.

Results

First, performance on the empirical measures of EF was examined, including developmental trends and relationships between the measures. Second, the results of the BRIEF-P were analyzed for both parent and teacher ratings, including factor analysis to examine the construct validity of the questionnaire. Third, correlational data was obtained to study the relationship between children's performance on the EF measures compared to their BRIEF-P ratings provided by parents and teachers.

Developmental Trends of Measures of Executive Functioning

Descriptive statistics on children's performance on the empirical executive functioning tasks are shown in Table 1.

Measures of Inhibition

Whisper Task. Older preschoolers performed significantly better on the Whisper task ($p = 0.006$) as shown in Table 1.

Gift Delay. The scores of the 3 scores (peeking score, number of peeks made, and latency to first peek) were highly correlated ($r_s \geq 0.85$) and therefore were standardized and

averaged to create one overall Gift Delay score. Older preschoolers performed significantly better than younger preschoolers on the Gift Delay task ($p = 0.047$) as shown in Table 1.

No significant correlations were found between the Whisper Task and the Gift Delay Task in either age group, even when sex and verbal ability were controlled for.

Tower of Hanoi

Older preschoolers performed significantly better on the Tower of Hanoi ($p < 0.001$) as shown in Table 1.

Working Memory

Backward & Forward Digit Span. Older preschoolers performed significantly better on the Backward Digit Span task ($p < 0.001$) and on the Forward Digit Span task ($p < 0.001$) as shown in Table 1. Backward Digit Span and Forward Digit Span performance were correlated only in the older children ($r = 0.492$, $n = 28$, $p < 0.01$).

Intradimensional-Extradimensional Shifting

Preliminary analyses indicated that the number of errors-to-criterion and the number of trials-to-criterion were highly correlated for all four phases of the task (r 's range from .79 to .95, $p < .01$) across both age groups. Further analyses, therefore, considered only the number of errors-to-criterion as a measure of task performance (Wall, 2003).

A 2-way repeated measures ANOVA with phase as the within subjects variable and the number of errors-to-criterion score for each of the 4 phases as the dependent variable revealed a significant main effect for phase ($F(3, 171) = 2.96$, $p = .034$) and for the interaction of phase and age ($F(3, 171) = 4.01$, $p = .009$), but not for the main effect of age ($F(1, 57) = 0.59$, $p = .444$). One-way MANOVA showed that there were no significant differences between the different age groups for the phases, except for the extradimensional phase ($F(1, 57) = 5.09$, $p = .028$). In this phase, older preschoolers ($M = 22.34$, $SD = 31.13$)

made more errors before reaching criterion compared to younger preschoolers ($M = 8.13$, $SD = 14.62$). Post-hoc tests revealed that the Intradimensional phase was easier ($M = 9.00$, $SD = 17.53$) compared to the compound reversal ($M = 17.90$, $SD = 24.17$) and ED phases ($M = 15.12$, $SD = 25.02$), and the compound discrimination ($M = 10.86$, $SD = 20.57$) was easier than the compound reversal. Table 1 shows these results with the errors-to-criteria expressed as standardized values.

Emotional Control/Regulation

Inter-rater reliability was calculated using interclass correlations for ratings provided by three raters when children received their first choice toy ($M = 0.81$, $p < 0.05$) and last choice toy ($M = 0.67$, $p < 0.052$). Interclass correlation for Anger ratings and Disgusted ratings were not included in calculations for the first-choice toy rating, nor were interclass correlations for Anger ratings as more than one rater provided the same ratings for all subjects for these scales.

A principal components factor analysis using oblimin rotation resulted in a 2-factor solution (Positive Affect, Negative Affect). This solution was used to score children's emotional displays after receiving a gift they found desirable (i.e., their first choice gift) and after receiving an undesirable gift (i.e., their last choice gift). A positive-affect measure was created by summing the ratings for each child on scores loading onto the Positive Affect factor: Engagement, Happy and Overall Positive Rating. The internal consistency across these dimensions was 0.77 for emotional displays after receiving the first choice gift and 0.61 for emotional displays after receiving the last choice gift. The negative-affect measure was created by summing the ratings for each child on scores loading on the Negative Affect factor: Disgusted and Anger. The internal consistency for Negative Affect was not

obtainable for the displays after receiving the first choice gift as all subjects were rated as having no anger displays. Internal consistency for the displays after receiving the last choice toy was 0.6 (see Table 2).

To obtain a measure emotion regulation each child's emotional-display score after the undesirable gift was subtracted from his or her emotional-display score after the desirable gift. Separate emotion-regulation scores were calculated for positive and negative affect and are shown in Table 1. There was no significant difference between age groups in the difference in positive ($p > 0.7$) or negative emotion regulation ($p > 0.1$).

PPVT-3

A one-way ANOVA revealed that there was a main effect for age for the PPVT-3 score ($F(1, 59) = 43.54, p < 0.001$). A 2-way (Age X Sex) ANOVA revealed a significant main effect for age ($F(1, 59) = 39.25, p < 0.001$), but no significant effects for the main effect for sex or the interaction between sex and age.

Relations Between of Measures of Executive Functioning

In both the younger and older preschoolers, relationships between the measures of executive functioning were fairly weak and few of these relationships were significant. However, somewhat stronger relations were observed in older preschoolers (see Table 3a & 3b).

BRIEF-P

Internal Consistency. Internal consistency was calculated using Cronbach's alpha (1951) and results are shown in Table 4. Generally, the results show excellent internal consistency. However, parent ratings for Emotional Control and Plan/Organize behavior were lower than the rest of the scales and internal consistency of the parent ratings was also lower than that of teachers. This finding is consistent with original internal consistency data provided by

the author's of the BRIEF-P (Gioia, Espy, & Isquith, 2003) who found a lower alpha for parent ratings on the Plan/Organize scale.

Reliability – Interrater Agreement. The correlations between the parent and teacher ratings were moderate, with an overall mean correlation of 0.27 (see Table 5). All parent and teacher ratings were significantly correlated (r s from 0.27 to 0.46) on all scales except Emotional Control ($r = 0.03$). Consistent with previous literature on parent and teacher ratings (BRIEF-P manual p.49; Offord, et al., 1996), parents rated children as having significantly greater problems on all scales with the exception of the Inhibition scale.

Effects of Sex. Two separate MANOVAs were conducted with age as the independent variable and the 5 scales of the BRIEF-P as the dependent variables for both the parent and teacher ratings. Results revealed that there was no difference on any of the scales on the parent ratings, but that teachers rated boys significantly higher on the Inhibition ($F(1, 58) = 8.96, p < 0.01$) and Shift scales ($F(1, 58) = 3.88, p = 0.054$).

Effects of Age. Two separate MANOVAs were conducted with sex as the independent variable and the 5 scales of the BRIEF-P as the dependent variables, one for parent ratings and one for teacher ratings. Results revealed that there was no difference in how parents or teachers rated the younger and the older children on any of the scales.

Factor Analysis. To examine the construct of the BRIEF-P, exploratory factor analyses were conducted on the parent ($N = 58$) and teacher ($N = 59$) ratings. Following the original analysis for the BRIEF-P (Isquith et al., 2004), principal factor analysis was used, with an oblique (Promax) rotation due to the likelihood of correlated factors. Since the *a priori* conceptual model of executive function is a multidimensional construct (Miyake et al., 2000), a single factor model was not considered. Furthermore, a four-factor model was not considered because, with only 5 scales, one factor would necessarily be defined by only one

scale. Analysis and the interpretation of the appropriate number of initial common factors was conducted based on eigenvalues, percentage of variance accounted for by the extracted factors, and scree plots (Isquith et al., 2004; Tabachnick & Fidell, 2001). In addition, subtests with factor loadings of $>.35$ were retained as measured indicators of the given factor.

For the parent sample, a two-factor model was retained as the best model in the analyses. The rotated factor loadings for this solution accounted for 75.5% of the variance (see Table 6). Examination of the solution indicated that three scales, Working Memory, Plan/Organize, and Inhibition loaded exclusively on the first factor. The Shift and Emotional Control scales loaded exclusively on the second factor. The two factors were correlated moderately ($r = 0.43$).

The same principal factor analysis was performed on the data from the teacher ratings and a two-factor solution similar to that in the parent sample was obtained, accounting for 78.3% of the variance. The rotated factor loadings for this solution are presented in Table 6. Two scales, Working Memory and Plan/Organize, loaded exclusively on the first factor and the Shift and Emotional Control scales defined the second factor. Both factors had a secondary loading of the Inhibition scale and the two factors were correlated moderately ($r = 0.48$).

Relationship Between EF Measures and BRIEF-P

EF Measures and Scales of the BRIEF-P. Performance on EF tasks was not strongly related to scales of the BRIEF-P for both parent and teacher reports, though some tasks were significantly related to scales of the BRIEF-P (see Table 7). Among these significant correlations, the EF measures that correlated with specific scales were often not the tasks that were intended to assess the component of EF the scale is intended to represent. For example, for parent ratings the Inhibit scale correlated significantly with children's

performance on the Extradimensional Shift of the IDED Shift task ($r = 0.28, n = 60, p < 0.05$) and with their change in Positive Affect in the Disappointment Procedure ($r = 0.26, n = 60, p < 0.05$). Conversely, teacher ratings resulted in no significant correlations between the Inhibit scale and measures of EF. In fact, there were no commonalities between parent and teacher ratings regarding significant relations between EF tasks and the scales of the BRIEF-P.

To further examine the relation between the scales of the BRIEF-P and the EF measures, scores on the empirical EF tasks were standardized and aggregated according to the primary construct they were intended to measure. The relations between these scores and scales of the BRIEF-P were fairly weak, though several significant correlations were found for parent ratings when age and sex were partialled out (see Table 8). For parent ratings, significant correlations were found between the EF measures of Shifting and the WM ($r = 0.32, n = 51, p < 0.05$) and Inhibit ($r = 0.29, n = 51, p < 0.05$) BRIEF-P scales, between the EF measures of Emotional Control and the Shift ($r = -0.31, n = 51, p < 0.05$) and Emotion Control ($r = -0.27, n = 51, p < 0.05$) BRIEF-P scales, and between the EF measures of Planning/Organizing and the Inhibit ($r = -0.29, n = 51, p < 0.05$) and WM ($r = -0.29, n = 51, p < 0.05$) BRIEF-P scales. For teacher ratings, significant correlations were found between EF measures of WM and the Planning/Organizing ($r = -0.27, n = 51, p < 0.05$) BRIEF-P scale, and between the EF measure of Emotional Control and the Shift ($r = -0.26, n = 51, p < 0.06$) BRIEF-P scale.

When parent and teacher ratings are combined, correlations between the EF tasks remain fairly weak and though some significant correlations are retained (see Table 9). The significant correlations that remained were between the performance on the Backward Digit Span and the Emotional Control ($r = -0.26, n = 57, p < 0.06$) and Planning/Organizing ($r =$

-0.27, $n = 57, p < 0.05$) scales of the BRIEF-P, between performance on the Forward Digit Span and the WM ($r = -0.29, n = 57, p < 0.05$) and Planning/Organizing ($r = -0.32, n = 57, p < 0.05$) scales of the BRIEF-P, between the score of Difference in Positive Affect and the Inhibit ($r = 0.27, n = 57, p < 0.05$), WM ($r = 0.27, n = 57, p < 0.05$) and Planning/Organizing ($r = 0.29, n = 57, p < 0.05$) scales of the BRIEF-P, and between the score of Difference in Negative Affect and the Shift ($r = -0.26, n = 57, p < 0.05$) and Emotional Control ($r = -0.33, n = 57, p < 0.05$) scales of the BRIEF-P.

EF Measures and Factors of the BRIEF-P. To further examine the relationship between the measures of EF and ratings on the BRIEF-P, factors were created from the parent and teacher ratings to reflect the factors established in the principle component factor analysis (see Table 10). That is, parent ratings were divided into 2 factors: an Emergent Metacognition factor that included ratings from all items assessing Working Memory, Planning/Organizing, and Inhibition and a Flexibility factor that included ratings assessing the Emotional Control and Shifting. Teacher ratings were also divided into 2 factors: an Emergent Metacognition factor identical to that for parent ratings, and a Flexibility factor that included ratings assessing Emotional Control, Shifting, and Inhibition.

The parent Emergent Metacognition factor was significantly related to EF measures that primarily assesses Planning/Organizing (Tower of Hanoi, $r = -0.27, n = 60, p < 0.05$), Working Memory (Forward Digit Span, $r = -0.27, n = 60, p < 0.05$), Shifting (Extradimensional shift, $r = -0.08, n = 60, p < 0.05$), and Emotional Control (Change in Positive Affect, $r = 0.29, n = 60, p < 0.05$; Change in Negative Affect, $r = -0.27, n = 60, p < 0.05$). The parent Flexibility factor was significantly related to a measure of Emotional Control (Change in Negative Affect, $r = -0.36, n = 60, p < 0.01$). Overall, the correlations between these two factors and children's performance on the EF measures were weak.

The teacher Emergent Metacognition factor was significantly related to an EF measure that primarily assesses Working Memory (Backward Digit Span, $r = -0.26$, $n = 60$, $p < 0.05$). The teacher Flexibility factor was also primarily related to an EF measure that primarily assesses Working Memory (Backward Digit Span, $r = -0.28$, $n = 60$, $p < 0.05$). As with the parent factors, the relations between the teacher factors and children's performance on the EF measures were weak.

Discussion

How do we measure a child's executive functioning abilities in a naturalistic setting? Although a small selection of studies have detailed the relative contribution of EF to children's everyday behaviors in a real-world context, they have done so in terms of children's social competence and problem behaviors. Until the development of the BRIEF and the BRIEF-P, researchers were limited to using developmentally appropriate tasks adapted from developmental cognitive neuroscience to assess children's EF. Examining EF via everyday behaviors can be considered an approach that is complementary (Isquith et al., 2004) to the typical individual assessments of EF using batteries of tasks. Although studies using the BRIEF and BRIEF-P have detailed the validity of the measurement in capturing various profiles of EF in distinct populations, it remained unclear how and whether children's performance on the BRIEF-P is related to their performance on traditional empirical EF tasks. The present study represents an attempt to clarify the ecological validity of various neuropsychological assessments of EF as well as examine the structural validity of the BRIEF-P.

Developmental Trends of Measures of EF

Significant developmental trends were found for children's performance on all measures of EF, with the exception of specific phases of the ID/ED shift task and both

measures of emotion regulation. These findings provide further support for the extensive amount of previous research which has shown that different components of EF have different developmental trajectories (e.g., Carlson, 2005; Luciana & Nelson, 1998; Hughes, 1998; Diamond & Taylor, 1996; Welsh et al., 1991). For example, compared to 3-year-olds, 6-year-old children solve more problems correctly on ToH tasks (Bull, Espy, & Senn, 2004). In addition, age related changes have been reported in EF measures of children's inhibitory control (Carlson & Moses, 2001), shifting (e.g., Carlson, Mandell, & Williams, 2004), and working memory (e.g., Gordon & Olson, 1998). Although age related changes of emotional control have been previously reported in the literature (e.g., Fox & Calkins, 2003), these studies used more traditional measures of emotional control such as coding of facial expression or physiological markers. The lack of age differences found in the present study is similar to findings of a recent study on which the Disappointment Procedure was based (Kieras, Tobin, Graziano, & Rothbart, 2005). Though the procedure used by Kieras et al. (2005) assesses the same construct using a slightly different paradigm, results of the study also showed no evidence that age (3- to 5-year-olds) significantly predicted children's displays of positive or negative affect. In addition, previous research using the ID/ED shift paradigm (e.g., Wall, 2003) did not find any ages differences between any of the 4 phases of the task, though the results were based on a small sample size and the ages tested ranged from 2- to 3-years. The present study provides further clarification regarding the difficulty of each of the phases in the ID/ED shift task, as well as the age-based differences on the performance of preschool children in each of the phases of the task.

More than one empirical EF task was used to measure the constructs of working memory and inhibition, and it was anticipated that children's performance on these measures, that evaluated the same construct, would be associated to some degree. The

absence of a relationship between measures, as found with the two measures of inhibition and the two measures of WM, may be attributed to the fact that each task assessed a specific component of the construct. A lack of correlation between the two measures of Inhibition (i.e., Gift Delay and Whisper tasks), may be attributed to the different aspects of inhibition the two tasks may primarily assess. The Gift Delay task can be regarded as measure of delay inhibition, as children have to delay the action of turning around and looking at the gift they are to receive. It can be said that the Whisper task may assess both WM and inhibition, as the task requires children to remember that they must whisper the names of the cartoon characters they are trying to identify. Following this reasoning, one would expect to find a correlation between the Whisper task and the measures of WM, the Forward and Backward digit span. The low correlations between these 3 tasks in both the younger and older age groups may be related to the accuracy of the two working memory tasks that measure the construct of WM. Evidence that casts doubt on the validity of the WM tasks comes from the finding that a significant correlation between the forward and backward digit span tasks was only found for the older preschoolers. Although both tasks have been used extensively to measure working memory in children and adults (e.g., Davis & Pratt, 1995), more developmentally appropriate tasks have been developed in recent years for use with children (e.g., Counting and Labeling; Carlson et al., 2002). The use of such a task in the present study would have perhaps yielded better results in terms of the relation to other tasks that may include a component of working memory.

The above findings lend further support to a recent suggestion by Carlson (2005) that the manner in which we fractionate EF does not necessarily map on directly to age changes in abilities or the difficulty levels of empirical tasks. Instead, the assessment of EF abilities through division into constructs (e.g., WM, Inhibition, and Shifting) may be most

useful in determining individual differences rather than specific abilities at certain ages.

Carlson et al. (2005) created difficulty scales for commonly used empirical EF tasks in order to gain clarity as to which tasks are most appropriate for a given age level or ability. These scales showed no clear pattern of certain tasks assessing a specific construct being easier or being passed earlier than others, however the most difficult tasks in all age groups studied (2 through 6 years) are those thought to require a combination of abilities (e.g., both WM and inhibition). Therefore, although the majority of previous studies, including the present one, advocate that certain task demands are dissociable based on their primary inherent demands (e.g., Espy & Bull, 2005; Hughes, 1998) or even their affective nature (i.e., hot vs. cool; Hongwanishkul et al., 2005; Zelazo & Müller, 2002) and focus on age related changes, future studies should investigate the patterns of relations between EF measures rather than their developmental sensitivity.

Relations Between EF Measures

The majority of studies involving the administration of a battery of tasks have examined the inter-task correlations via factor analysis (see Zelazo & Müller, 2002 for review), which resulted in several key factors of EF. The absence of strong significant correlations between the measures of EF in the current study is inconsistent with previous studies in terms of the strength of the correlations (e.g., Carlson & Moses, 2001; Hughes, 1998) and may be due to the range of ages tested and a smaller sample size. In previous studies, the range of age groups assessed is typically narrower and the sample size within these age groups larger (e.g., Carlson & Moses, 2001). The older children tested in the present study may have demonstrated more consistency in their performance resulting in a lack of differentiation between their scores of the various EF tasks. If older children were to be tested, their more established personality and more pronounced cognitive style may lead

them to more differentiated performances on the various EF tasks. The absence of significant correlations between some tasks was not unexpected as past research has shown there some EF measures are indeed not related to each other. For example, Welsh et al. (1999) have shown that performance on the TOH is not related to measures of WM that assess capacity (e.g., span tasks) nor to certain measures of inhibition (e.g., resistance to interference in a Stroop task).

Construct Validity of the BRIEF

In the present study, ratings on the BRIEF-P by parents and teachers confirm previously reported results by the authors of the questionnaire (Isquith et al., 2004), with the exception of fewer factors were obtained for the current data. The absence of the third factor, Inhibitory Self-control, may be due to the fact that factors of planning/organizing and inhibition are sufficiently accounted for by a more general factor such as Emergent Metacognition. Also, inhibition may be sufficiently intertwined with the Emergent Metacognition factor and the Flexibility factor, as shown by its secondary loading onto both factors in the teacher sample.

No significant age or sex results were found with the exception of teacher ratings for boys on inhibitory control and shifting abilities. Previously published results with the BRIEF-P also showed similar patterns with respect to boys' performance as rated by teachers, but overall, the original (Isquith et al, 2004) study revealed small but significant age differences. This variation in findings of age between the current and the study by Isquith and colleagues (2004) may be attributed to the difference in age groups used: in the study by Isquith and colleagues, the only age differences found were for 3-year-olds compared to 2-, 4-, and 5-year-olds on their self-inhibitory abilities. If the current study had included a younger age range, it is possible that age differences would have been detected. Both studies

found no age differences for older preschool children (i.e., above 3-year-old), which calls into question the ability of the BRIEF-P to discriminate levels of EF in terms of everyday behaviors past a certain age. Isquith et al. (2004) do recognize that the BRIEF-P provides a means for examining multiple domains of EF in a consistent manner with the caveat that the ratings “represent less differentiated, more intertwined abilities than what is observed in school-age children.” In the developmental literature, empirical studies have suggested that EF abilities may be undifferentiated in the early preschool years and become more modular as children develop (Carlson, 2005; Zelazo & Müller, 2002), though some components of EF such as Working Memory and Inhibition, are known to develop earlier (Espy et al., 2004). The lack of age differences may also be attributed to the use of mean dichotomization in the present study to establish the two age groups. Creating the groups based on the number of children assessed and their respective chronological ages may have led to a loss of power to be able to detect age differences in ratings on the BRIEF-P.

Overall, the BRIEF-P may be limited in its sensitivity in capturing the early stages of EF development in a typical population as it was designed for use with a clinical population. The items and the rating scale may only be appropriate in evaluating extreme cases of executive dysfunction in preschool children. The minor variations in EF abilities seen between children in a typical population may not appear when using this questionnaire as the range of scores produced between subjects would be similar. The present study found similar mean scores for the 5 scales of the BRIEF-P as rated by parents and teachers to those found in the original BRIEF-P (Gioia et al., 2003): ratings for both age groups in the present study were within one standard deviation from the mean of those reported in the BRIEF-P manual. In addition, the variability among the scores of the 5 scales based on parent and teacher ratings are such that only small significant differences were found

between 2- to 3-year-olds and 4- to 5-year old by the authors of the questionnaire (Gioia et al., 2003). Therefore, differences in BRIEF-P scores may not be large enough to be detected and limit the interpretation of the scores provided by raters. As a result, although the BRIEF-P is the first assessment available for examining EF abilities in preschool children in terms of their everyday behaviors it seems to be best suited for neuropsychological assessments of dysfunction in cognitive abilities rather than evaluation of typical development of EF.

Relationships between the BRIEF-P and measures of EF

Based on the low and scarce significant correlations between the EF tasks and the ratings of parents and teachers on BRIEF-P (both scales and factors), it is apparent that the questionnaire and the empirical tasks are measuring the primary constructs of EF in distinct manners that may not be comparable. Although the results of this study confirm that the BRIEF-P possesses a reliable set of internally consistency scales that tap EF (Isquith et al., 2004), it cannot be said that it accurately gauges the relative components that contribute to EF in preschoolers as measured by empirical tasks. It may be that the BRIEF-P assesses a general, more metacognitive construct of EF that primarily includes WM, Inhibition, and Planning/Organizing skills in terms of children's everyday behaviors. If so, the BRIEF-P should only be used to determine a general level of EF in children and not to identify deficits in any of the scales. An alternate interpretation is that the BRIEF-P items are more specific than the measures of EF, as they were designed to tap into particular behaviors that are associated with high frontal lobe functioning (Anderson et al., 2002). In contrast, the cognitive measures may be thought of as more "multi-determined" (Anderson, 1998) and require a range of lower-order cognitive skills while still assessing one primary construct of EF.

The absence of strong significant correlations between the EF measures and the BRIEF-P also sheds light on the ecological validity of tasks widely used in developmental psychology to assess EF in preschool children. If the BRIEF-P is considered an appropriate measure of everyday behaviors, then a conclusion to be drawn from the current study is that results of traditional laboratory EF tasks do not reflect children's abilities in everyday situations. Although no direct claim has been made by researchers that these measures are highly related to children's performance in a naturalistic setting, it is evident from these results that the presence of such an association is questionable as only weak correlations between laboratory EF tasks and the BRIEF-P were found. Similar conclusions were arrived at in a study of children with traumatic brain injury who were administered the BRIEF and individual empirical tests of EF (Vriezen & Pigott, 2002), as no correlations were found between the two types of assessment. A concern may also be raised with regards to the empirical tasks themselves. Although not directly addressed above, the low reliability of the empirical EF tasks may be the cause of low correlations found between the BRIEF-P ratings and performance on the EF tasks.

An obvious explanation for the limited number and weak significant relations between results from the BRIEF-P and the EF tasks is the differential "sampling frames" (Vriezen & Pigott, 2002) used for the two types of measures of EF in children. Parents and teacher's assessment via the BRIEF-P are based on observations they have made of a child within a real-world setting, while the neuropsychological measures are done in a more structured, predictable environment. Overall, it is quite clear the BRIEF-P measures different aspects of EF compared to empirically based individual tests of EF used with preschool children. The use of both parent and teacher ratings also may contribute to the poor relations between the BRIEF-P and the EF measures as utilizing two raters presents an

issue of differential sampling frames within the BRIEF-P itself. Parents and teachers each employ different points of reference when rating children on behaviors questionnaires such as the BRIEF-P: teachers compare, perhaps unconsciously, the child in question's behavior to other children under their supervision while parents have a more limited pool of comparison subjects that may include the child's siblings or other family members. In addition, children's temperament may influence the way they are rated by both teachers and parent: their perceived disposition may alter the way that raters evaluate their performance by exaggerating problems or underrating difficulties. As a result, parents and teacher apply different norms when rating children's behaviors, with teachers using many more children as their anchors for evaluation. The results of the present study clearly emphasize these differences, but also highlight the possibility that multiple formats of assessing EF in preschool children may be needed depending on the nature of research goals. Future research with regards to the effects of children's disposition, such as temperament, in relation to the BRIEF-P performance would be warranted and easily accomplished with the administration of a questionnaire such as the Children's Behavioral Questionnaire (Rothbart et al., 2001).

The types of statistical analyses conducted in the present study may have imposed a limitation by influencing the final outcome regarding the relation between children's performance on the EF tasks and their ratings on the BRIEF-P. Data collected from both children and parents were handled using a manifest variable approach in that the variables being examined were thought to be observable and therefore measured directly. Though this type of approach is widely used in studies examining EF development in children (e.g., Carlson et al., 2004), there has been a recent movement advocating the use of latent variable models where variables are thought to describe an unobservable construct that cannot be

measured directly. Latent-variable analysis is thought to be helpful when examining executive functioning as the analysis statistically extracts the common variance between the tasks that are chosen to tap the different aspects of EF, separating the variance due to executive processes from the considerable variance due to nonexecutive task requirements and measurement error (Miyake et al., 2001). It is thought that the field is in need of systematic studies that use measurement model approaches, such as the latent variable approach, to better characterize executive function organization in children, similar to what was initiated by Miyake et al. (2000) in college students (Espy et al., 2004). Structural equation modeling (SEM) is an example of a type of measurement model approach that could be used, as is path analysis which is thought to have properties similar to SEM: “both involve modeling complex relations among variables, comparing complicated models to simpler, nested models, and conducting multiple-group comparisons to determine whether a model fits the observed data from different groups equally well” (Senn et al., 2004, p. 458). Path analysis is considered a powerful tool by which to investigate relations among executive functions or other neuropsychological abilities. Evidence that stems from path analysis must be considered preliminary (Senn et al., 2004), however, as a limitation is that only one test can be selected as the primary measure for each construct being evaluated (e.g., Backward Digit Span for WM). Although tests can be selected to primarily measure a given construct on the basis of empirical evidence and theoretical rationale, the instruments may not fully assess the executive process being targeted (Senn et al., 1984). The use of SEM analysis in the present study may have allowed for a better assessment of the latent constructs of by using the shared variance between the multiple empirically based tasks believed to assess EF (Senn, et al., 2004).

The next phase of research investigating children's EF abilities in a real-world context would be to determine what methods, in addition to parental and teacher rating scales, could be used to assess EF through children's everyday behaviors. There is the possibility of modifying traditional empirical tasks to assess components of EF that based more on real-world experiences child may encounter in their everyday lives, resulting in more quasi-naturalistic tasks. Examples of such efforts can be seen in the area of the development of emotion regulation (Hrabok, Kerns, & Müller, in prep; Kieras et al., 2005) and inhibition (Carlson et al., 2005). In the present study, the Disappointment procedure may be regarded as a quasi-naturalistic task as children may be presented with gifts they do not like in their natural environments. However, since strong significant correlations were not found between performances as evaluated by both change in positive and negative affect and the Emotional Control scale of the BRIEF-P, perhaps the use of quasi-experimental tasks is not sufficient. An alternative would be to combine traditional experimental tasks with both quasi-experimental tasks and naturalistic observations to assess EF in manners that are relatable to children's everyday behaviors in various contexts. The development of such studies is necessary for the extension of empirical work to areas outside of the world of laboratory research.

Conclusion

The analysis of preschoolers' EF abilities via the BRIEF-P in conjunction with traditional empirical tasks is an initial step in understanding the development of EF and how it relates to children's everyday behaviors. Although reliable parent and teacher reports of children's behavioral manifestations of EF abilities may have a high level of ecological validity in understanding children's abilities in the real-world, there are limitations to using such methods (Isquith et al., 2004). Behavior ratings scales such as the BRIEF-P may

provide simply a general or global view of EF abilities of children within a real-world setting, and therefore would need to be considered a complementary tool in EF assessment alongside neuropsychological based tests that are more developmentally sensitive. In general, the BRIEF-P should be considered by researchers as a measure of EF abilities at the domain-general level, and may be optimal for use in studies that examine individual differences rather than age related changes.

References

- Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological Rehabilitation, 8*, 319-349.
- Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology, 8*, 231-240.
- Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology (Special Issue: Working Memory), 49A*, 5-28.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions. *Psychological Bulletin, 121*, 65-94.
- Baron, I.S. (2000). Test Review: Behavior Rating of Executive Function. *Child Neuropsychology, 6*, 235-238.
- Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition, 55*, 30-40.
- Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*, 48-71.
- Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo & L.G. Tassinary (Eds.). *Handbook of psychophysiology (2nd ed.)*. (pp. 602-642). New York, NY, US: Cambridge University Press.
- Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology, 26*, 571-593.
- Bull, R., Espy, K. A.; Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology & Psychiatry, 45*, 743-754.
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273-293.
- Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbit (Ed.), *Methodology of frontal and executive function* (pp. 81-116). Hove: Psychology Press.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*, 595-616

- Carlson, S. M., Davis, A. C., & Leach, J. G. (2005). Less is More: Executive Function and Symbolic Representation in Preschool Children. *Psychological Science, 16*, 609-616.
- Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and Prediction From Ages 2 to 3. *Developmental Psychology, 40*, 1105-1122.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032-1053.
- Carlson, S. M., & Moses, L. J., & Breton, C. (2002).. How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73-92.
- Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology, 87*, 299-319.
- Carvajal, H. H., Parks, J. P., Logan, R. A., & Page, G. L. (1992). Comparisons of the IQ and vocabulary scores on Wechsler Preschool and Primary Scale of Intelligence—Revised and Peabody Picture Vocabulary Test-Revised. *Psychology in the Schools, 29*, 22-24.
- Case, R. (1992). The role of the frontal lobes in the regulation of cognitive development. *Brain and Cognition, 20*, 51-73.
- Cole, P.M., Martin, S. E., & Dennis, T.A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development, 75*, 317-333.
- Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science, 7*, 547-552.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- das Gupta, P., & Bryant, P. E. (1989). Young children's causal inferences. *Child Development, 60*, 1138-1146.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450-466.
- Davis, H.L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*, 25-31.
- Dempster, F. N. (1991). Inhibitory processes: A neglected dimension of intelligence. *Intelligence, 15*, 157-173.

- Diamond, A. (1991). Developmental time course in human infants and infant monkeys, and the neural bases of inhibitory control reaching. *Annals of the New York Academy of Sciences: Part VII. Inhibition and executive control*, 637-704.
- Diamond, A., Kirkham, N. Z., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, 38, 352-362.
- 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, 315-334.
- Donders, J. (2002). The Behavior Rating Inventory of Executive Function. *Child Neuropsychology*, 8, 229-230.
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, 33, 261-268.
- Dunn, L., & Dunn, L. (1997). *The Peabody Picture Vocabulary Test* (3rd edition). Circle Pines, MN: American Guidance Service.
- Espy, K. A., & Bull, R. (2005). Inhibitory processes in young children and individual variation in short-term memory. *Developmental Neuropsychology*, 28, 669-688.
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematics skills in preschool children. *Developmental Neuropsychology*, 26, 465-486.
- Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology*, 24, 51-73.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483-527.
- Fox, N., & Calkins, S. D. (2003). The development of self-control of emotion: Intrinsic and extrinsic influences. *Motivation and Emotion*, 27, 7-26.
- Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, 39, 3-27.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½-7 years old on a Stroop-like day-night test. *Cognition*, 53, 129-153.
- Gioia, G. A., Espy, K. A., & Isquith, P.K. (2003). *The Behavior Rating Inventory of Executive Function – Preschool Version*. FL: Psychological Assessment

Resources.

- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *The Behavior Rating Inventory of Executive Function*. Lutz, FL: Psychological Assessment Resources.
- Gioia, G. A., Isquith, P. K., Kenworthy, L., & Barton, R. (2002). Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology*, 8, 121-137.
- Gordon, A. C. L., & Olson, D. (1998). The relationship between acquisition of a theory of mind and information processing capacity. *Journal of Experimental Child Psychology*, 68, 70-83.
- Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, 8, 26-32.
- Hodapp, A. F. (1993). Correlation between Stanford-Binet IV and PPVT-R scores for young children. *Psychological Reports*, 73, 1152-1154.
- Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, 28, 617-644.
- Hrabok, M., Kerns, K.A., & Müller, U. (in prep). Cognitive and emotional dimensions of self-regulation in 4-year-old children.
- Hudson, J. A., Sosa, B. B., & Shapiro, L. R. (1997). Scripts and plans: The development of preschool children's event knowledge and event planning. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 77-102). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology*, 34, 1326-1339.
- Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health*, 7, 131-142.
- Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for central executive dysfunction in autism. *Neuropsychologia*, 32, 477-492.
- Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, 26, 403-422.
- Kagan, J. (1994). On the nature of emotion. *Monographs of the Society for Research in Child Development*, 59, (2-3, Serial No. 240): 7-24.
- Kieras, J.E., Tobin, R. M., Graziano, W. G., & Rothbart, M. K. (2005). You Can't

Always Get What You Want: Effortful Control and Children's Responses to Undesirable Gifts. *Psychological Science*, 16, 391-396

- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6, 449-467.
- Konchanska, G., Murray, K. T., Jacques, T. Y., Koenig, A. L., & Vandegest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, 67, 490-507.
- Lehto, J. (1996). Working memory capacity and summarizing skills in ninth-graders. *Scandinavian Journal of Psychology*, 37, 84-92.
- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *The British Journal of Developmental Psychology*, 21, 59-80.
- Levin, H. S., Culhane, K. A., Hartmann, J., Evankovich, K., Mattson, S. J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J. M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology*, 7, 377-395.
- Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, 36, 273-293.
- Mangelsdorf, S. C., Shapiro, J. R., & Marzolf, D. (1995). Developmental and temperamental differences in emotional regulation in infancy. *Child Development*, 66, 1817-1828.
- 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.
- Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.). *Unsolved mysteries of the mind* (p. 93-148). Hove: Erlbaum.
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111-121.
- Nigg, J. T. (2001). Is ADHD a disinhibitory disorder? *Psychological Bulletin*, 127, 571-598.
- Offord, D. R., Boyle, M. H., & Racine, Y. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child & Adolescent Psychiatry*, 35, 1078-1085.

- Paris, S. G., & Newman, R. S. (1990). Developmental aspects of self-regulated learning. *Educational Psychologist, 25*, 87-102.
- Perner, J., & Lang, B. (1999). Development of theory of mind and cognitive control. *Trends in Cognitive Science, 3*, 337-344.
- Peskin, J., & Ardino, V. (2003). Representing the mental world in children's social behavior: Playing hide-and-seek and keeping a secret. *Social Development, 12*, 496-512.
- Robbins, T. W. (1998). Dissociating executive functions of the prefrontal cortex. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.). *The prefrontal cortex: Executive and cognitive functions* (p. 117-130). Oxford: Oxford University Press.
- Roberts, R.J. Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology, 12*, 105-126.
- 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*, 1394-1408.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.
- Scholnick, E. K., Friedman, S. L., & Wallner-Allen, K. L. (1997). What do they really measure? A comparative analysis of planning tasks. In S. L. Friedman & E. K. Scholnick (Eds.), *The developmental psychology of planning: Why, how, and when do we plan?* (pp. 127-156). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using Path Analysis to Understand Executive Function Organization in Preschool Children. *Developmental Neuropsychology, 26*, 445-464.
- Stuss, D. T., & Knight, R. T. Eds, (2002). *Principles of frontal lobe function*, Oxford: Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th edition.). Boston, MA: Allyn & Bacon.
- Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology, 8*, 296-303.
- Wall, P. (2003). *The Development of Reversal Learning and Extra-Dimensional Shifting in Preschoolers*. Undergraduate Thesis, Department of Psychology, University of Toronto.

- Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, 7, 131-149.
- Wolfe, C. D., & Bell, M. A. (2003). Working memory and inhibitory control in early childhood: contributions from physiology, temperament, and language. *Developmental Psychobiology*, 44, 68-83.
- Zelazo, P. (1999). Language, levels of consciousness, and the development of intentional action. In P. D. Zelazo, J. W. Astington & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self control* (pp. 95-118). Mahwah, NJ: Lawrence Erlbaum Associates.
- Zelazo, P. D. (2000). Self-reflection and the development of consciously controlled processing. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 169-189). Hove, East Sussex: Psychology Press Ltd.
- Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, 1, 198-226.
- Zelazo, P. D., & Müller, U. (2002). Executive functions in typical and atypical development. In U. Goswami (Ed.), *Handbook of Childhood Cognitive Development* (pp. 445-469). Oxford: Blackwell.
- Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The Development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, Serial 264 (68).
- Zelazo, P.D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representations. *Developmental Psychology*, 35, 1059-1071.

Appendix A

Table 1
Descriptive Statistics & Performance on All EF Measures as a Function of Age

Measure	Younger preschoolers (<i>n</i> =30)	Older Preschoolers (<i>n</i> =30)	Age effects	Range
<i>Inhibition Measures</i>				
Whisper	1.67 (0.40)	1.90 (0.19)	<i>t</i> (58) = -2.84**	0-2
Gift Delay	-0.09 (0.41)	0.11 (0.32)	<i>t</i> (55) = -2.03*	standardized
Tower of Hanoi	6.77 (5.03)	16.4 (10.52)	<i>t</i> (58) = -4.52**	0-27
<i>Working Memory Measures</i>				
Backward Digit Span	0.10 (0.31)	0.93 (0.83)	<i>t</i> (57) = -5.07**	2-4
Forward Digit Span	2.03 (0.98)	3.30 (0.95)	<i>t</i> (57) = -5.03**	0-8
<i>ID/ED Shifting</i>				
Compound Discrimination	0.20 (1.23)	-0.20 (0.67)	<i>t</i> (57) = 1.63	standardized
Compound Reversal	-0.04 (1.02)	0.04 (0.99)	<i>t</i> (57) = -0.21	standardized
Intradimensional Shift	-0.14 (0.60)	0.15 (1.29)	<i>t</i> (57) = -1.13	standardized
Extradimensional Shift	-0.28 (0.58)	0.29 (1.24)	<i>t</i> (57) = -2.26*	standardized
<i>Disappointment Procedure</i>				
Difference in Positive Affect	2.83 (2.32)	2.62 (1.78)	<i>t</i> (57) = 0.40	- 12 to 30
Difference in Negative Affect	-0.67 (1.45)	-0.17 (0.60)	<i>t</i> (57) = -1.70	- 8 to 20

Note. **p* < 0.05, two-tailed;

***p* < 0.01, two-tailed

Table 2
Means of Emotion Ratings for Children's Emotional Regulation

Affect Rating	After First Choice (Desirable) Toy		After Last Choice (Undesirable) Toy	
	Younger	Older	Younger	Older
Engaged	4.03 (0.56)	3.93 (0.59)	3.87 (0.51)	3.93 (0.59)
Happy	3.13 (0.90)	3.21 (1.05)	2.17 (0.91)	2.38 (0.86)
Overall Positive Rating	3.30 (0.99)	3.48 (0.88)	1.60 (0.93)	1.69 (0.93)
Positive Affect Composite (sum)	10.47 (2.21)	10.62 (1.99)	7.63 (1.79)	8.00 (1.83)
Anger	1.00 (0)	1.00 (0)	1.17 (0.38)	1.03 (0.19)
Disgusted	1.00 (0)	1.03 (0.19)	1.50 (1.14)	1.17 (0.47)
Negative Affect Composite (sum)	2.00 (0)	2.03 (0.19)	2.67 (1.45)	2.21 (0.56)

Note. N = 60. Standard deviations are in parentheses.

Table 3a & 3b
Correlations between EF measures

Younger Preschoolers												
EF Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. Whisper	-	0.00	0.23	0.06	-0.02	-0.08	0.05	-0.14	-0.02	0.19	-0.28	0.00
2. Gift Delay		-	0.16	0.23	-0.22	0.23	-0.02	-0.04	-0.02	0.38*	0.05	0.48**
3. Tower of Hanoi			-	0.21	0.34	0.02	0.07	-0.16	-0.06	0.08	0.06	0.26
4. Backward Digit Span				-	0.34	0.15	-0.06	0.04	0.08	0.21	-0.24	0.31
5. Forward Digit Span					-	-0.05	-0.29	-0.15	-0.35	-0.19	-0.06	0.11
6. Compound Discrimination						-	0.31	-0.07	0.12	0.21	0.04	-0.25
7. Compound Reversal							-	0.18	0.43*	0.13	-0.13	-0.32
8. Intradimensional Shift								-	0.60**	0.04	0.22	-0.04
9. Extradimensional Shift									-	-0.01	0.17	-0.14
10. Difference in Positive Affect										-	-0.29	0.22
11. Difference in Negative Affect											-	-0.13
12. PPVT												-
Older Preschoolers												
EF Measures	1	2	3	4	5	6	7	8	9	10	11	12
1. Whisper	-	-0.14	-0.17	-0.09	0.07	0.10	0.00	-0.20	-0.01	0.16	-0.26	0.07
2. Gift Delay		-	0.35	0.30	0.06	-0.14	0.12	0.12	0.01	0.11	0.53**	-0.01
3. Tower of Hanoi			-	0.30	0.14	-0.04	0.10	0.07	-0.32	0.21	0.09	0.34
4. Backward Digit Span				-	0.49**	-0.03	0.37^	-0.14	-0.06	0.15	-0.01	0.64**
5. Forward Digit Span					-	0.33	0.25	0.00	0.05	-0.18	-0.16	0.34
6. Compound Discrimination						-	0.43*	0.54**	0.32	0.04	-0.29	-0.06
7. Compound Reversal							-	0.45*	0.16	-0.08	0.11	-0.02
8. Intradimensional Shift								-	0.62**	-0.35	0.10	-0.16
9. Extradimensional Shift									-	-0.27	0.12	0.02
10. Difference in Positive Affect										-	-0.11	0.11
11. Difference in Negative Affect											-	-0.40*
12. PPVT												-

Note. ^p < 0.06, two-tailed; *p < 0.05, two-tailed; **p < 0.01, two-tailed

Table 4
Internal Consistency (α) for Parent and Teacher Ratings on BRIEF-P

Scale	Parent ¹	Teacher ²
Inhibit	0.85	0.96
Shift	0.82	0.83
Emotional Control	0.71	0.92
Working Memory	0.90	0.95
Plan and Organize	0.65	0.91

Note. ¹ $n=58$. ² $n=59$.

Table 5
Correlations Between BRIEF-P Parent and Teacher Ratings

Parent Ratings	Teacher Ratings			
	Inhibit	Shift	Emotional Control	Working Memory
Inhibit	0.29*			
Shift		0.32*		
Emotional Control			0.03	
Working Memory				0.46**
Plan/Organize				0.27*

Note. $N = 58$

* $p < 0.05$, two-tailed; ** $p < 0.01$, two-tailed

Table 6
Factor Loadings Three-Factor Model for Parent and Teacher Ratings on BRIEF-P

Scale	Parent Sample		Teacher Sample	
	Factor 1	Factor 2	Factor 1	Factor 2
Inhibit	0.92		0.57	0.41
Shift		0.93		0.77
Emotional Control		0.80		0.91
Working Memory	0.86		0.99	
Plan and Organize	0.78		0.99	
Factor correlations (<i>r</i>)				
Factor 2	0.43		0.48	
Cumulative % of variance	75.5%		78.3%	

Table 7
Correlations between EF measures and BRIEF-P Parent and Teacher Ratings

EF Measures	Parent Ratings					Teacher Ratings				
	Inhibit	Shift	Emotional Control	Working Memory	Plan/Organize	Inhibit	Shift	Emotional Control	Working Memory	Plan/Organize
1. Whisper	0.22	-0.04	0.02	0.00	0.09	-0.16	0.03	-0.16	-0.07	-0.02
2. Gift Delay	-0.24	-0.15	-0.10	-0.06	-0.14	-0.14	-0.10	-0.22	0.06	-0.04
3. Tower of Hanoi	-0.24	-0.13	-0.31	-0.24	-0.30	-0.11	0.05	-0.18	-0.16	-0.13
4. Backward Digit Span	-0.07	0.02	-0.08	-0.11	-0.15	-0.21	-0.19	-0.29*	-0.24	-0.28*
5. Forward Digit Span	-0.15	0.12	0.05	-0.28*	-0.30*	-0.11	-0.09	-0.12	-0.23	-0.20
6. Compound Discrimination	-0.02	0.19	0.04	0.09	0.12	-0.02	0.12	0.02	0.16	0.10
7. Compound Reversal	-0.10	0.22	0.16	0.19	0.23	-0.11	0.18	0.07	0.00	-0.05
8. Intradimensional Shift	0.12	0.10	0.10	0.13	0.07	-0.16	0.04	-0.08	0.02	-0.06
9. Extradimensional Shift	0.28*	0.01	0.10	0.36**	0.18	-0.09	-0.06	-0.17	0.07	0.03
10. Difference in Positive Affect	0.26*	-0.12	-0.02	0.24	0.27*	0.19	-0.13	0.21	0.21	0.21
11. Difference in Negative Affect	-0.18	-0.27*	-0.36**	-0.29*	-0.23	0.01	-0.15	-0.11	-0.03	-0.06
12. PPVT	-0.17	-0.04	-0.20	-0.20	-0.24	-0.22	-0.21	-0.27*	-0.17	-0.14

Note. * $p < 0.05$, two-tailed; ** $p < 0.01$, two-tailed

Correlations are between standardized scores of EF measures and raw scores on BRIEF-P.

Table 8
Partial Correlations between EF Constructs and BRIEF-P Parent and Teacher Ratings

EF Constructs	Parent Ratings					Teacher Ratings				
	Inhibit	Shift	Emotional Control	Working Memory	Plan/Organize	Inhibit	Shift	Emotional Control	Working Memory	Plan/Organize
1. Inhibition	-0.03	-0.12	-0.03	-0.05	0.02	-0.15	0.02	-0.21	-0.04	-0.03
2. Shift	0.13	0.14	0.02	.32*	.29*	-0.01	0.16	0.01	0.21	0.14
3. Working Memory	-0.25	0.09	0.04	-0.24	-0.23	-0.11	-0.12	-0.15	-0.26	-.27*
4. Emotional Control	0.03	-.31*	-.27*	-0.06	0.02	0.12	-0.26^	-0.04	0.15	0.07
5. Plan/Organize	-.29*	0.02	-0.13	-.29*	-0.22	-0.05	0.13	-0.04	-0.15	-0.10

Note. ^ $p < 0.06$, two-tailed; * $p < 0.05$, two-tailed; ** $p < 0.01$, two-tailed

Partial correlations control for age and sex.

For EF Constructs, standardized scores were aggregated in following manner:

Inhibition = Whisper & Gift Delay Task

Shift = Compound Discrimination, Compound Reversal, Intradimensional Shift, & Extradimensional Shift

Working Memory = Backward Digit Span & Forward Digit Span

Emotional Control = Change in Positive Affect & Change in Negative Affect

Plan/Organize = Tower of Hanoi

Table 9
Correlations between EF measures and BRIEF-P Ratings

EF Measures	Parent & Teacher Ratings				
	Inhibit	Shift	Emotional Control	Working Memory	Plan/Organize
1. Whisper	-0.07	0.01	-0.19	-0.03	0.002
2. Gift Delay	-0.2	-0.21	-0.15	-0.04	-0.11
3. Tower of Hanoi	-0.17	0.02	-0.24	-0.24	-0.23
4. Backward Digit Span	-0.19	-0.1	-0.26 [^]	-0.22	-0.27*
5. Forward Digit Span	-0.16	0.04	-0.10	-0.29*	-0.32*
6. Compound Discrimination	-0.03	0.20	0.04	0.15	0.14
7. Compound Reversal	-0.13	0.24	0.16	0.09	0.1
8. Intradimensional Shift	-0.06	0.08	0.01	0.07	-0.003
9. Extradimensional Shift	0.06	-0.03	-0.06	0.24	0.12
10. Difference in Positive Affect	0.27*	-0.14	0.09	0.27*	0.29*
11. Difference in Negative Affect	-0.08	-0.26*	-0.33*	-0.17	-0.17
12. PPVT	-0.25 [^]	-0.14	-0.38**	-0.21	-0.24

Note. [^] $p < 0.06$; * $p < 0.05$, two-tailed; ** $p < 0.01$, two-tailed

Table 10
Correlations between EF measures and BRIEF-P Parent and Teacher Factors

EF Measures	Parent Ratings		Teacher Ratings	
	Emergent Metacognition	Flexibility	Emergent Metacognition	Flexibility
1. Whisper	0.07	-0.04	-0.11	-0.15
2. Gift Delay	-0.19	-0.15	-0.05	-0.19
3. Tower of Hanoi	-0.27*	-0.10	-0.15	-0.12
4. Backward Digit Span	-0.12	-0.03	-0.26*	-0.28*
5. Forward Digit Span	-0.27*	0.10	-0.19	-0.13
6. Compound Discrimination	0.07	0.13	0.08	0.02
7. Compound Reversal	0.11	0.22	-0.07	0.00
8. Intradimensional Shift	0.13	0.11	-0.08	-0.12
9. Extradimensional Shift	-0.08*	0.06	0.00	-0.13
10. Difference in Positive Affect	0.29*	-0.08	0.22	0.16
11. Difference in Negative Affect	-0.27*	-0.36**	-0.02	-0.07
12. PPVT	-0.23	-0.14	-0.21	-0.29*

Note. * $p < 0.05$, two-tailed; ** $p < 0.01$, two-tailed

Parent Emergent Metacognition = Working Memory + Plan/Organize + Inhibition

Parent Flexibility = Emotional Control + Shift

Teacher Emergent Metacognition = Working Memory + Plan/Organize + Inhibition

Teacher Flexibility = Emotional Control + Shift + Inhibition