Factor Structure and Utility of the Behavior Rating Inventory of Executive Function-Preschool Version

Jamie A. Spiegel,
Department of Psychology and the Florida Center for Reading Research, Florida State University.

Christopher J. Lonigan, and
Department of Psychology and the Florida Center for Reading Research, Florida State University.

Beth M. Phillips
Department of Educational Psychology and Learning Systems and the Florida Center for Reading Research, Florida State University.

Abstract

Executive function (EF) is a domain general cognitive construct associated with a number of important developmental outcomes. The *Behavior Rating Inventory of Executive Function-Preschool Version* (BRIEF-P) is intended to assess five distinct components of EF in preschool age children. In this study, a series of factor analyses were conducted with teacher-reported EF of 2,367 preschool students to assess the structure of the BRIEF-P, and the predictive relations between the resulting factors and children's academic abilities and behavioral self-regulation were assessed to test the construct and convergent validity of the BRIEF-P scores. Results yielded mixed findings concerning the structure of the BRIEF-P and validity of its resultant scores. Results of the factor analyses indicated that the items of the BRIEF-P did not map onto factors in the way that would be expected based on its item-to-subscale mapping. The best solutions were a four-factor and a bifactor model. The four-factor solution revealed substantial correlations between factors, and although the bifactor solution identified a general Self-Regulation factor that explained variance in responses across items, this general factor did not account for all of the overlap among specific factors. Analyses of the relations for the factors from the correlated-factors and the bifactor models indicated that the majority of the factors had limited convergent validity with academic ability or with a measure of behavior self-regulation. Overall, these findings call into question the validity of aspects of BRIEF-P.

Keywords

Self-regulation; Executive Function; prekindergarten/preschool; emergent literacy; psychometrics; factor analysis

Early self-regulation skills have pervasive and long lasting implications for the development of children's academic abilities and psychopathology (e.g., McClelland et al., 2007;

Correspondence regarding this article should be directed to Jamie A. Spiegel (spiegel@psy.fsu.edu) or Christopher J. Lonigan (lonigan@psy.fsu.edu), Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee FL 32306-4301.

Matthews, Ponitz, Morrison, 2009; Duncan et al., 2007; Vitaro, Brendgen, Larose, & Tremblay, 2005). The construct of self-regulation is broadly defined in the literature as consisting of multiple aspects of children's behavior, including planning, controlling, and directing abilities. From a cognitive perspective, the construct most frequently associated with self-regulation is executive function (EF). EF is a domain general cognitive process, associated with the self-regulation of behavior and emotion, as well as the development of academic abilities and social skills.

Although a variety of skills have been proposed as part of the construct of EF, including problem solving, delayed responding, attention selectivity, flexibility, and goal setting and planning (e.g., Mahone & Hoffman, 2007), results of factor analytic studies with adults (e.g., Miyake, Friedman, Emerson, Witzki, & Howerter, 2000) and older children (e.g., Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003) indicate that EF is best characterized by three inter-correlated factors: inhibitory control, working memory, and shifting. Inhibitory control is defined as the ability to inhibit a prepotent response in favor of a subordinate response. Working memory is defined as the ability to maintain, update, and manipulate information within memory. Shifting is defined as the ability to switch attention between mental sets or tasks, or the ability to engage and disengage with specific aspects within tasks (Miyake et al., 2000). These EF skills can be assessed as early as infancy (Diamond, 1985, 1991). Recent conceptualizations of EF posit that substantial developments in EF skills occur throughout the first six years of life, with marked changes occurring between ages 3 and 6 (e.g., Diamond & Taylor, 1996; Espy, Kaufmann, McDiarmid, & Glisky, 1999) and between ages 7 and 9 (Anderson, 2002).

Strong relations between self-regulation and other developmentally important constructs, including academic and behavioral outcomes, have consistently been reported from concurrent and longitudinal data (e.g., Allan, Hume, Allan, Farrington, & Lonigan, 2014; Best, Miller, Naglieri, 2011, Isquith, Gioia, & Espy, 2004; McClelland et al., 2007). Specifically, results of research have demonstrated substantial relations between self-regulation and academic outcomes, including literacy, mathematics, and school readiness, in young children (e.g., Allan & Lonigan, 2011; Blair & Razza, 2007; McClelland et al., 2007). More specific associations have also been examined, with findings indicating unique relations between self-regulation and pre-literacy skills, including phonological awareness, print knowledge, and vocabulary skills (e.g., Fuhs, Farran, & Nesbitt, 2014; Sims & Lonigan, 2013). Findings tend to indicate the largest associations between EF skills and orthographic knowledge and phonological awareness (i.e., Shaul & Schwartz, 2013; Sims & Lonigan, 2013). The strength of these relations may vary based on the type of self-regulation measure being implemented (i.e., performance vs. report-based ratings; Fuhs, Farran, & Nesbitt, 2014) and the specific self-regulation skills being assessed (e.g., Jacob & Parkinson, 2015). Findings from a recent meta-analysis indicate that the mean effects of EF skills, specifically IC, on academic outcomes in young children vary by method of EF assessment, with cool EF performance-based tasks (i.e., tasks which do not involve punishment or reward for performance) and teacher reports producing the largest effects (Allan et al., 2014). These findings suggest that when examining the relation between EF skills and academic ability, the preferred methods of assessment are either cool EF performance tasks or teacher reports.

Improving the accuracy with which EF skills are measured has been proposed as a means through which the understanding of the complex relations between cognition, development, and environment can be improved (Espy, 2004). Although improvements in performance measures of EF for preschool-age children have been made in recent years, a criticism of these assessments is that they have limited ecological validity (Bodnar, Prahme, Cutting, Denckla, & Mahone, 2007). Performance tasks assess EF skills under situationally constrained, highly standardized conditions rather than the real world applications of such skills.

Rating scales involving either self-report or parent and teacher reports of EF skills in everyday situations have been developed in an effort to address the issue of ecological validity. Although performance tasks and ratings scales are designed to assess the same underlying constructs, a recent comparison of commonly used EF performance tasks and rating scales demonstrated only small to moderate associations between these two types of EF measures (Toplak, West & Stanovich, 2013). Toplak et al. hypothesized that EF performance tasks and rating scales measure different aspects or levels of the same construct. Specifically, they suggested that performance tasks provide information about EF skills under optimal, externally constrained situations, whereas rating scales provide information about EF skills in every-day, goal-directed situations.

A commonly used rating-scale measure of EF in children is the *Behavior Rating Inventory of Executive Function* (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000). The BRIEF is a parent- or teacher-rating scale of everyday EF behaviors in school-age children. The BRIEF was designed to be used in conjunction with performance-based measures to allow for a comprehensive understanding of children's EF skills. Similar to Toplak et al. (2013), Gioia et al. suggested that performance tasks be used to assess more specific components of EF, whereas the BRIEF is designed to assess broader EF skills and their everyday, real-world applications. The BRIEF has been used to examine the relation between EF skills and important developmental outcomes, including academic ability (e.g., Mcauley, Chen, Goos, Schachar, & Crosbie, 2010) and psychopathologies (e.g., Toplak, Bucciarelli, Jain, & Tannock, 2008). Gioia, Espy, and Isquith (2003) adapted the BRIEF for use with preschool-age children. The resulting measure, the *Behavior Rating Inventory of Executive Function-Preschool Version* (BRIEF-P), consists of 63 Likert-scale items that make up five clinical scales and three indices. The BRIEF-P is intended to be completed by teachers or parents of 2- to 6-year-old children, with ratings based on the child's observed behaviors within the relevant environments (i.e., home, school).

The initial scale-development process for the BRIEF-P involved adjusting items from the BRIEF to better reflect the preschool context, developing new items to reflect preschool-specific behaviors, assigning these items to scales, and removing items with high means, wide dispersion indices, and low correlations based on item-total correlations (Isquith et al., 2004). A series of exploratory factor analyses (EFA) with oblique rotations were then conducted to help clarify the scale structure of the BRIEF-P, using both parent ($N = 460$) and teacher ($N = 302$) ratings of preschool children. In the first EFA, 63 items were used as indicators, and the analysis yielded five factors: Inhibit, Shift, Emotional Control, Working Memory, and Plan/Organize. Isquith et al. did not report item-level factor loadings; however,

they reported that the scales had strong internal consistency, with values ranging from .80 (Plan/Organize) to .90 (Inhibit) for parent ratings and from .90 (Shift) to .97 (Plan/Organize) for teacher ratings. Isquith et al. also conducted scale-level EFAs on parent and teacher ratings, which yielded a three-factor model: Inhibitory Self-Control (comprised of the Emotional Control and Inhibit scales), Flexibility (comprised of the Shift and Emotional Control scales), and Emergent Metacognition (comprised of the Working Memory and Plan/Organize scales). The three-factor solution accounted for 87% of the variance in the parent-report data and 92% of the variance in the teacher-report data.

Although the BRIEF-P is one of the few EF rating scales appropriate for use with preschool-age children, few studies have been conducted to assess its structure. To our knowledge, only three studies since Isquith et al. (2004) have attempted to validate the factor structure of the BRIEF-P (Bonillo et al., 2012; Duku & Vaillancourt, 2014; Ezpeleta, Granero, Penelo, Osa, & Doménech, 2012). Of these studies only Duku & Vaillancourt (2014) and Ezpeleta et al. (2012) examined the convergent validity of the resultant factors with other measures of EF. Across both studies, small-to-moderate associations between the BRIEF-P subscales and other report-based measures of EF were found, and Ezpeleta et al. (2012) found no significant associations between BRIEF-P scales and a performance-based measure of EF. No study to date has examined the association of the BRIEF-P with other developmental outcomes.

Bonillo et al. (2012) translated the BRIEF-P into Catalan and collected parent and teacher reports for a sample of 400 Catalan-speaking 3- to 6-year-old children from North East Spain. Results from this study demonstrated that the five scales of the Catalan version of the BRIEF-P had high internal consistency, comparable to that of the English version as reported by Isquith et al. (2004). Bonillo et al. then conducted an item-level confirmatory factor analysis (CFA) of the BRIEF-P. Their five-factor model failed to converge because the correlations between the factors were consistently greater than 1.0 (i.e., a Heywood case), indicating that fewer dimensions underlie the BRIEF-P than were proposed by Isquith et al. Because the five-factor model proposed by Isquith et al. did not converge, Bonillo et al. evaluated a unidimensional model; however, the unidimentional model yielded poor model fit.

Ezpeleta et al. (2012) also conducted an item-level CFA of the Catalan version of the BRIEF-P using teacher reports on 620 3-year-old children from Barcelona, Spain. They reported that a hierarchical model (i.e., 5 first-order factors and 3 second-order factors) provided an acceptable fit to the data. However, this model did not include four items (i.e., "overreacts to small problems," "has to be more closely supervised than similar playmates," "has trouble changing activities," "has trouble remembering something, even after a brief period of time") that had very high levels of endorsement. Except for the exclusion of these four items, the results of this analysis offered support for the five-factor model of the BRIEF-P proposed by Isquith et al. (2004).

Finally, Duku and Vaillancourt (2014) reported a series of factor analyses of the BRIEF-P using parent and teacher reports on 625 typically developing preschool-age children from Canada. They first attempted to replicate the second-order factor solution reported by Isquith

et al. (2004); however, the model was not identified. They then conducted item-level CFAs to compare a five-factor and a one-factor model. Neither the five-factor model nor the one-factor model provided adequate fit to the data. In a third series of analyses, Duku and Vaillancourt conducted categorical CFAs to test the unidimensionality of the model for each scale. Results of these analyses indicated that for both parent and teacher reports, the Emotional Control, Plan/Organize, and Working Memory scales met the criteria for unidimensionality; however, the Inhibit and Shift scales did not, suggesting that more than five dimensions underlie the BRIEF-P. Duku and Vaillancourt then conducted a categorical EFA, and results of this analysis indicated that an eight-factor model provided the best fit to the data. In this solution, the original Inhibit scale was divided into two scales with cross-loading items (Awareness and Impulsivity scales), and the original Shift scale was divided into three new scales with cross-loading items (Inflexibility, Adjusting, and Sensory scales). This eight-factor model varied by informant, such that different items loaded onto the scales for parents' reports versus teachers' reports. Internal consistency for these scales ranged from below adequate (i.e., $\alpha = .64$ for the Adjusting scale) to high (i.e., $\alpha = .94$ for the Impulsivity scale). A final CFA was conducted integrating the five new scales with the original Emotional Control, Plan/Organize, and Working Memory scales. Fit statistics for this analysis indicated acceptable fit across parent and teacher informants.

Current Study

The purpose of the current study was to examine the factor structure of the BRIEF-P as proposed by Isquith et al. (2004) and to assess the construct and convergent validity of the scores on the resultant factors with academic ability and performance-based measures of EF. The equivocal nature of the findings regarding the factor structure of the BRIEF-P calls in to question the validity with which these relations can be assessed using this measure, as it is currently unclear the extent to which individual scales from the BRIEF-P represent the underlying constructs they purportedly measure. The inability to replicate the factor structure of the BRIEF-P at the item-level indicates that the factor structure reported by Isquith et al. may not be the best structure for this measure. Therefore, we hypothesized that the factor structure proposed by Isquith et al. would not provide adequate fit to the data and that alternate models would need to be assessed.

Given the limited research to date on the convergent validity of scores on the BRIEF-P subscales and academic ability, our hypotheses about these relations were largely based on associations consistently reported between other measures of EF and academic outcomes (e.g., Allan et al., 2014; Best et al., 2011; Clark, Pritchard, & Woodward, 2010). Specifically, findings that indicate unique relations between specific EF skills and academic abilities (e.g., Allan & Lonigan, 2011; Jacob & Parkinson, 2015; Sims & Lonigan, 2013) and school readiness (e.g., Zevenbergen & Ryan, 2010), led us to hypothesize that the BRIEF-P factors would have substantial and unique associations with children's performance on measures of early academic skills. Furthermore, although the relations between the measures of academic ability and school readiness used in the current study and the BRIEF-P were not previously evaluated, these academic measures were chosen based on prior research findings indicating substantial relations with other measures of self-regulation. Specifically, findings from prior research have consistently indicated unique

associations between self-regulation and different emergent literacy skills (e.g., Allan & Lonigan, 2011; Sims & Lonigan, 2013), as measured by the Test of Preschool Early Literacy (Lonigan, Wagner, Torgesen, & Rashotte, 2007) and with school readiness (e.g, Zevenbergen & Ryan, 2010), as measured by the Bracken School Readiness Assessment (Bracken, 2002).

Similarly, to our knowledge, limited research to date has been published on the associations between the BRIEF-P and performance-based measures of EF; however, findings from the available research (i.e., Kraybill & Bell, 2013) indicate substantial associations between these methods of measurement. Further, significant relations between the Head-Toes-Knees-Shoulders task (HTKS; Ponitz et al., 2008) and other report-based ratings of self-regulation (e.g., Ponitz et al., 2009) have been reported. Multiple studies using the HTKS have shown that it essentially functions as an inhibitory control task (Matthew et al., 2009; Ponitz et al., 2009; Wanless et al., 2011), although McClelland argues that it measures several constructs related to self-regulation. Therefore, we hypothesized that the factors of the BRIEF-P would be strongly correlated with HTKS scores.

Method

Participants

Children for this study were recruited for a larger study designed to examine the impacts of early childhood curricula on the school-readiness skills of preschool children at risk of educational difficulties. Data for this study came from 2,367 children in 109 classrooms who were recruited from private preschools, public preschools, and Head Start centers in Massachusetts (40% of sample) and New Mexico (60% of sample). Recruited preschools served primarily children from high poverty families or served children with identified developmental difficulties (e.g., language delay). Children in the study ranged in age from 29 to 74 months ($M = 53.02$, $SD = 6.66$), and slightly more than half of the sample (55%) were boys. The sample was racially and ethnically diverse: 27.9 percent were White; 1.5 percent were Black/African American; 36.5 percent were Latino; 2.7 percent were Asian; 1.4 percent were Native American; .5 percent were multi-racial; and 29% were not reported/ unknown. The sample used in this study was more racially and ethnically diverse than the sample used by Isquith et al. (2004), which was predominantly white (i.e., ~70%). As a group, children scored in the low-average range on standardized measures of language and early literacy skills (i.e., average standard scores of 86 to 93).

Measures

Behavior Rating Inventory of Executive Function-Preschool Version—Teachers completed the BRIEF-P (Isquith et al., 2004). The BRIEF-P was designed to assess day-to-day EF skills as rated by parents or teachers. It consists of 63-items and is composed of five scales: Inhibit, Shift, Emotional Control, Working Memory, and Plan/Organize. Items are rated on a three-point scale (i.e., never, sometimes, often), with higher scores representing lower EF. Internal consistency reliabilities for these scales range from .80 (Plan/Organize) to .90 (Inhibit), and the subtests have moderate to high test-retest reliability over the course of 4.5 weeks (e.g., rs = .78-.90; Gioia et al., 2003).

Test of Preschool Early Literacy—Children were administered the *Test of Preschool Early Literacy* (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), which includes three subtests: Definitional Vocabulary, Phonological Awareness, and Print Knowledge. The Definitional Vocabulary subtest includes 36 items, and each item includes two parts. For the first part of each item, the child had to label a single image or group of images that she or he is shown, and for the second part of the item, the child had to respond to a follow-up question regarding function or relevant context for the item. The second part of each item enables the measure to assess a more definitional, depth of vocabulary dimension in addition to the simple confrontational naming task.

The Phonological Awareness subtest includes 27 items that requires the child to either blend sounds together to form a word (blending; e.g., blending "star" and "fish" to form "starfish," blending /f/ and "ox" to form "fox") or to remove sounds from a word to make a new word (elision; e.g., removing "snow" from "snowshoe" to create "shoe," removing /d/ from "raid" to create "ray"). Answers for all blending and elision items are real words, and items spanned the range of linguistic complexity (e.g., compound words to phonemes). There are both multiple-choice items, which require children to point to one picture out of four that represents the correct response, and free-response items, which require children to verbally produce the correct response. The blending component includes six multiple-choice and nine free-response items, and the elision component includes six multiple-choice and six free-response items. The Print Knowledge subtest consists of 36-items that assess early print concepts, alphabet recognition, letter-name knowledge, and letter-sound knowledge. Internal consistency reliabilities for these subtests are high for 3-, 4-, and 5-year-old children (i.e., αs of .89 to .95), and the subtests have moderate to high validity correlations with other measures of similar constructs (e.g., *r*s = .58 – 77; Lonigan et al., 2007).

Bracken School Readiness Assessment (BSRA)—Children's school readiness was assessed using the *Bracken School Readiness Assessment* (BSRA; Bracken, 2002). In the BSRA children are assessed on six basic skills typically taught to children to prepare them for formal education (i.e., colors, letters, numbers and counting, sizes comparisons, shapes). The BSRA is designed for use with children in preschool through second grade. The BSRA has been shown to have strong psychometric characteristics (Panter & Bracken, 2009). The BSRA is an adapted version of the Bracken Basic Concepts Scale (Bracken, 1984); which has been shown to be reliable for use with 3-, 4-, 5-, and 6-year-old children (i.e., α of .73 to .93; Bracken, 1987).

Head-Toes-Knees-Shoulders Task—Children's behavioral self-regulation was assessed using the HTKS (McClelland et al., 2007; Ponitz et al., 2008). On the HTKS, children are asked to inhibit a dominant response and respond to commands in a conflicting, non-automatic manner. Specifically, they are told to remember four rules: if the administrator says 1) "touch your head" the correct response is to touch one's toes, 2) "touch your toes" the correct response is to touch one's head, 3) "touch your knees" the correct response is to touch one's shoulders, and 4) "touch your shoulders" the correct response is to touch one's knees. Children are given two practice trials and then 20 scored trials. For the first 10 trials, children are only tested on the "head" and "toes" rules. In the second 10 trials,

children are tested on all four rules. No points are awarded if the child touches the wrong body part, two points are awarded if the child touches the correct body part, and one point is awarded if the child moves in the direction of the wrong body part, but is able to self-correct and ends by touching the correct body part. Consequently, the maximum possible score was 40. Internal consistency for the HTKS is high (α = .93), and scores on the HTKS correlate with teachers' ratings of classroom behavior (Ponitz, McClelland, Matthews, & Morrison, 2009) and other direct measures of self-regulation (e.g., Allan & Lonigan, 2011; 2014).

Procedure

Prior to data collection, parents/guardians of children provided written informed consent/ permission for their children's participation in the project. All data for this study were collected in the fall of children's preschool year. Once consent was obtained, trained research assistants completed the battery of pre-academic and EF measures with the children across multiple testing sessions lasting from 15 to 40 minutes, depending on the schedule of the preschool and the child's attentional capacity. Commensurate with direct assessment of children, classroom teachers were asked to complete the BRIEF-P on the consented children in her or his classroom.

Not all children completed all measures during the fall assessment. Of the 2,367 children with teacher-completed BRIEF-P data, completed pre-academic assessments were obtained on between 89 percent (n = 2,100) and 75 percent (n = 1,774) of the children, depending on the specific assessment, and 17 percent (n = 392) of the children completed the HTKS. Data were missing for the pre-academic assessments because of child absences; the HTKS was added to the original assessment battery after pretesting was completed for most children. Children with and without completed data did not differ on sex, age, or BRIEF-P total scores. However, children missing TOPEL subscales, BSRA, or HTKS outcomes were more likely to be missing ethnicity/race information. Additionally, children missing HTKS data tended to have lower academic ability and school readiness compared to children who completed the HTKS.

Results

Identifying the Factor Structure of the BRIEF-P

Because items on the BRIEF-P are categorical, all factor analyses were conducted in Mplus (Muthén & Muthén, 2011) using the weighted least squares means and variance adjusted estimator (WLSMV). The WLSMV is a robust estimator that does not require variables to be normally distributed and is considered the best estimation option for modeling categorical or ordered data (Brown, 2006). The distribution of responses to items on the BRIEF-P were highly skewed, with most children rated at the low end of the rating scale for each item, and only 3 to 14 percent of children rated at the high end of the rating scale across all items. Less than one percent of the item-level data was missing across children. Full-information-maximum-likelihood estimation was used to account for the missing data in all analyses. All confirmatory and structural models were estimated using a sandwich estimator to account for the nested structure of the data (i.e., children were nested in classrooms).

Model fit was assessed using several fit indices. Although a nonsignificant χ^2 value indicates that a model provides an excellent fit to the data, models with many items and large samples rarely meet this threshold. Additional indices used to assess model fit, included the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized (or weighted) root mean square residual (SRMR or WRMR). In general, CFI and TLI values greater than .90 suggest adequate fit, and CFI and TLI values greater than .95 suggest good fit. RMSEA values of .05 and below suggest good fit; RMSEA values of .08 suggest moderate fit; and RMSEA values greater than .10 suggest poor fit. Lower values of SRMR and WRMR indicate better model fit (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). Nested model comparisons were conducted using the χ^2 difference test through the "DIFTEST" function in Mplus, with a nonsignificant χ^2 difference interpreted as support for the more parsimonious model.

We first attempted to fit the five-factor model that corresponded to the item-to-subscale mapping of the BRIEF-P reported by Isquith et al. (2004). Although this model converged, one item (Item 35) had an out-of-range standardized factor loading (i.e., loading > 1.0). Once this item was excluded, the model converged and all remaining items had acceptable factor loadings. The fit of this model was just less than adequate based on CFI and RMSEA values (see Model 1 in Table 1); however, it provided a better fit to the data than did a one-factor model (see Model 2 in Table 1), χ^2 ($N = 2,453$, $df = 2$) = 405.25, $p < .001$.

Because the model corresponding to the item-to-scale mapping of the BRIEF-P did not yield an acceptable solution when using all items and it did not provide an adequate fit to the data when using the items for which acceptable parameters could be computed, we next conducted an exploratory two-level factor analysis that examined up to five factors at both within (child) and between (classroom) levels. Examination of model fit statistics for models with more than one between-level factor revealed that adding more than one between-level factor generally resulted in decreases in overall model fit across different within-level models. Model fit indices for the one- to five-factor models are shown in Table 1 (see Models 3 – 7) for models with one between-level factor (all items loaded substantially on the between-level factor, with loadings ranging from .63 - .96, all ps < .001). These fit indices indicate that the three-, four-, and five-factor models provided good fits, with the five-factor model providing the best fit to the data (because these models are not nested and because WLSMV estimation in Mplus does not yield information-criteria statistics, these models could not be compared statistically). Inspection of the factor loadings for the five-factor model revealed that no item had its primary loading on the fifth factor, and the highest loading for any item on the fifth factor was .30. Therefore, the four-factor solution was determined to provide the best fit to the data with items having substantive factor loadings on each factor. Rotated (oblique) factor loadings for the four-factor solution from the exploratory two-level factor analysis are shown in Table 2.

Next, we used the results from the exploratory analyses to fit a four-factor model in a confirmatory factor analysis. In the first model, each item was allowed to load only on the factor with which the item had its highest loading. This model provided a less than adequate fit to the data (see Model 8 in Table 1). Consequently, we allowed the 17 items with factor

loadings equal to or greater than .30 on a second factor in the exploratory factor analysis (see Table 2) to crossload on the additional factor. This model provided an adequate fit to the data (see Model 9 in Table 1), and the model with crossloaded items fit the data significantly better than the model without crossloaded items, χ^2 ($N = 2,453$, $df = 5$) = 469.95, $p < .001$. Factor loadings for this model are shown on the left side of Table 3, and the correlations between factors are shown in the upper panel of Table 4. Factor 1 was composed primarily of items from the BRIEF-P Working Memory and Plan/Organize subscales (Factor $1_{WM/PO}$). Factor 2 was composed primarily of items from the BRIEF-P Shifting subscale (Factor 2_{SH}). Factor 3 was composed primarily of items from the BRIEF-P Emotional Control subscale (Factor 3_{EC}), and Factor 4 was composed primarily of items from BRIEF-P the Inhibit subscale (Factor 4_{IH}).

Finally, because of the substantial correlations between factors in the four-factor model, we fit a bifactor model with one general and four specific factors to determine if a common influence on items could account for the correlations between factors. The bifactor model provided an adequate fit to the data (see Model 10 in Table 1), and a comparison of the correlated-factors 4-factor model with the bifactor model yielded a significant chi-square difference test, χ^2 ($N = 2,453$, $df = 18$) = 399.74, $p < .001$, indicating that the bifactor model provided a better fit to the data than did the correlated-factors model; however, the model modification indices revealed that constraining the correlations between factors to zero in the bifactor model resulted in some model misspecification, indicating that the general factor did not account for all of the overlap between the specific factors. Factor loadings for the bifactor model are shown on the right side of Table 3.

Relations of BRIEF-P with Measures of Academic Skills and Behavioral Self-Regulation

Correlations of the latent variables from the correlated-factors and bifactor models with observed scores on the HTKS and measures of language skills, early literacy skills, and basic concepts are shown in the bottom panel of Table 4. As can be seen in the table, scores on these measures were most highly correlated with Factor $1_{WM/PO}$ and least correlated with Factor 3_{EC} of the correlated-factors model. The correlations of Factor $1_{WM/PO}$ and the measures of academic skills and HTKS were significantly stronger than those of Factors 2_{SH}, 3_{EC}, and 4_{IH} (by Steiger's [1980] T2 statistic, all ps < .001). The correlations between Factor 2_{SH} and the measures of academic skills and HTKS scores were significantly stronger than those between Factor 3_{EC} and all measures (all ps < .001), and they were significantly stronger than those of Factor 4_{IH} with Definitional Vocabulary and HTKS (ps < .01). Finally, Factor 4_{IH} was more strongly correlated with all measures than was Factor 3_{EC} (all ps < .05). Structural models in which each of the measures was simultaneously regressed on all of the factors revealed that only Factor $1_{WM/PO}$ accounted for significant unique variance. Although parameters for other terms were significant in some models, the fact that their relations with the measures reversed in sign relative to their zero-order correlations indicates that they were acting as suppressor variables in the structural models (see Table 4).

A similar pattern to that obtained with the factors from the correlated-factors model was evident for the correlations between the specific factors of the bifactor model and the measures of academic skills and HTKS. Factor $1_{WM/PO}$ was significantly correlated with all

measures. The correlations for Factors 2_{SH}, 3_{EC}, and 4_{IH} were uniformly low, and most were nonsignificant, with the exception of the correlations between these factors and Definitional Vocabulary. Additionally, correlations between the general factor and the measures of academic skills and HTKS were significant. Because all factors are orthogonal to each other in the bifactor model, the zero-order correlations indicate the unique variance accounted for by each factor in the academic measures and the HTKS.

To determine whether the associations between the BRIEF-P factors with the academic outcomes were substantially different for the subset of children who also had HTKS data the above analyses were replicated using only cases in with HTKS data were available. In these analyses, the pattern of results was similar those reported in Table 4, with moderate variations in some parameter estimates. The results of these additional analyses demonstrated that the pattern of results reported in Table 4 on the full sample hold for the subsample of children not missing the HTKS. Details from these analyses can be found in Table S1 in online supplemental materials.

Finally, we compared the strength of the correlations between the factors from the BRIEF-P and the academic measures to the strength of the correlations between HTKS scores and the academic measures (rs = .20, .33, 37, and .32 for vocabulary, print knowledge, phonological awareness, and Bracken scores, respectively). For the correlated-factors model, except for the fact that the correlations of HTKS and Factor 2_{SH} with Definitional Vocabulary were not statistically different, HTKS scores had stronger correlations with the academic measures than did Factors 2_{SH}, 3_{EC}, and 4_{IH} (by Steiger's T2 statistic, all ps < .001). HTKS scores correlated more strongly with Print Knowledge than did Factor $1_{WM/PO}$ (p < .05), but Factor $1_{WM/PO}$ correlated more strongly with Definitional Vocabulary than did HTKS scores (p < .001). For the bifactor model, the correlations of the HTKS with all academic measures except Definitional Vocabulary were stronger than those of the general factor, and Factors $1_{WM/PO}$, 2_{SH}, 3_{EC}, and 4_{IH}. The general factor, Factor 4_{IH}, and HTKS scores were equally correlated with Definitional Vocabulary, and Factor $1_{WM/PO}$ correlated more strongly with Definitional Vocabulary than did HTKS scores (p < .001). The strength of the correlations between the HTKS and academic outcomes also was compared to the strength of the BRIEF-P factors with academic outcomes within the subset of children with available HTKS data. Results from these analyses replicated those found in the whole sample and indicated that, in most cases, the correlations between the HTKS and academic outcomes were stronger than the correlations between the BRIEF-P factors and the academic outcomes (see Table S1).

Discussion

The primary purposes of this study were to determine whether the factor structure of the BRIEF-P proposed by Isquith et al. (2004) would be obtained in teacher reports of EF, in a large sample of preschool-age children, and whether the relations between obtained factors and measures of academic skills and a direct measure of self-regulation matched theorized relations. The five-factor model proposed by Isquith et al. did not yield an adequate fit to our data; therefore, a series of factor analyses were conducted to identify a better-fitting model for the BRIEF-P. A two-level EFA indicated that a facture structure with four within-level

factors and one between-level factor was the best solution. However, this four-factor model yielded less than adequate model fit when items were not allowed to crossload. When 17 items were allowed to load onto more than one factor, adequate model fit was achieved. A bifactor solution of the four-factor correlated-factors model yielded better fit than did the correlated-factors model, although inclusion of the general factor did not account for all of the covariation among the specific factors. Analyses that examined the associations between the factors from the correlated-factors and bifactor models with academic and performance-based EF yielded mixed results concerning the convergent and construct validity of the BRIEF-P factors. Overall, these results call into question the adequacy of the BRIEF-P for assessing specific facets of preschool children's EF.

Structure of the BRIEF-P

Although the BRIEF-P was constructed to measure five EF-related skills, factor analyses of teacher reports in this study did not support this factor structure. For the five-factor model to yield an acceptable solution, Item 35 (i.e., difficulties switching activities) had to be removed. Ezpeleta et al. (2012) had similar difficulties with their item-level analysis of the BRIEF-P and had to remove four items, including Item 35, for the model to converge. However, unlike the findings of Ezpeleta et al., in these data, a model that included all of the items with acceptable parameters provided inadequate model fit; indicating that the five-factor model proposed by Isquith et al. (2004) did not adequately account for the covariances among items of the BRIEF-P.

To identify the optimal solution for the data, we used exploratory two-level factor analyses in which the number of factors at the classroom and child levels was allowed to vary from one to five; a solution with one classroom-level and four-child level factors yielded good model fit, and all items had substantive loadings on at least one factor. These results indicate that although teachers varied overall in how they rated children in their classrooms on BRIEF-P items (i.e., tending to rate all children higher or lower), this variance was consistent across items. For the child-level factors, items from the factors Isquith et al. (2004) labeled as Inhibit, Emotional Control, and Shifting tended to coalesce into similar factors. The items from the factors Isquith et al. labeled Working Memory and Plan/Organize tended to coalesce into a single factor, similar to the Emergent Metacognition scale proposed by Isquith et al. A CFA of this four-factor model yielded inadequate model fit; however, a CFA of this four-factor model that allowed 17 items to cross-load provided acceptable model fit.

Moderate to high correlations between BRIEF-P factors have consistently been reported (e.g., Bonillo et al., 2012; Duku & Vaillancourt, 2014; Ezpeleta et al., 2012). In this study, correlations between multiple factors in the four-factor solution were moderate to large (e.g., $r = .60$ between Factors 3_{EC} and 4_{IH}, $r = .72$ between Factors $1_{WM/PO}$ and 4_{IH}), indicating substantial overlap in the constructs measured by these factors. To determine if these substantial correlations reflected the influence of some general construct across items, we examined a bifactor model that was based on the four-factor model, including the 17 crossloaded items. Although the bifactor model provided a better fit to the data than did the four-factor model, inclusion of a general factor did not eliminate the correlations between

the specific factors. Therefore, although some of the variance in items was common--indicating the presence of a general Self-Regulation factor, the substantial correlations were not solely the result of a common influence across items.

In both the correlated-factors model and the bifactor model, the factor structure of the BRIEF-P has some resemblance to relations between the five subscales and the three composite indices proposed by Isquith et al. (2004; see Sherman & Brooks, 2010, Figure 1). In this model of the BRIEF-P, the Working Memory and Plan/Organize subscales combine to represent the Emergent Metacognition Index; the Inhibit and Emotional Control subscales combine to represent the Inhibitory Self-Control Index; and the Shifting and Emotional Control subscales combine to represent the Flexibility Index. In the correlated-factors model from on our analyses, items primarily from the Working Memory and Plan/Organize subscales formed Factor $1_{WM/PO}$. Items primarily from the Shifting subscale formed Factor 2_{SH}, and items primarily from the Emotional Control subscale formed Factor 3_{EC}. Factors 2_{SH} and 3_{EC} had the highest number of items with crossloadings (i.e., almost all items from the Emotional Control subscale had substantive loadings on both Factor 2_{SH} and Factor 3_{EC}), suggesting the influence of the construct intended to be represented by the BRIEF-P Flexibility Index. Items primarily from the Inhibit subscale formed Factor 4_{IH}. There was a substantial correlation between Factor 3_{EC} and Factor 4_{IH}, suggesting the emergence of the construct represented by the BRIEF-P Inhibitory Self-Control Index. However, the high correlations between Factor $1_{WM/PO}$ and Factors 2_{SH} and 4_{IH}, which were higher than the within-index correlations, were inconsistent with the model proposed by Isquith et al. (2004).

Predictive Utility of BRIEF-P Factors

Analyses of the relations between our BRIEF-P factors and measures of early academic skills and behavioral self-regulation indicated that parts of the BRIEF-P could be useful for understanding the influences associated with higher or lower skill development. Based on the consistent associations reported between measures of EF and academic outcomes in samples of preschool children (e.g., Allan & Lonigan, 2011; Blair & Razza, 2007; Fuhs et al., 2014; McClelland et al., 2007), we had hypothesized that the BRIEF-P factors would have substantial and unique associations with children's performance on measures of early academic skills. Similarly, because results of previous research have revealed substantial relations between the HTKS task and report-based ratings of self-regulation (e.g., Ponitz et al., 2009), we had hypothesized that the factors of the BRIEF-P would be strongly correlated with HTKS scores.

Consistent with these hypothesized relations, zero-order correlations between academic outcomes and the BRIEF-P factors from the correlated-factors model were statistically significant, albeit small to moderate in size, for Factor $1_{WM/PO}$, Factor 2_{SH}, and Factor 4_{IH}. Correlations of HTKS scores with Factor $1_{WM/PO}$ and Factor 2_{SH} were also statistically significant. Factor 3_{EC} was not significantly correlated with HTKS scores or any academic skill measure, and HTKS scores were uncorrelated with Factor 4_{IH}. Although modest in size, at best, these relations were similar in magnitude to those reported in prior studies

comparing report-based and performance-based EF skills with young children (e.g., Eisenberg et al., 2000; Smith-Donald, Raver, Hayes, & Richardson, 2007).

Correlations between these academic and performance-based EF measures and the BRIEF-P factors from the bifactor model were less consistent with the hypothesized relations. In the bifactor model, only the general Self-Regulation factor and Factor $1_{WM/PO}$ (and Factor 2_{SH} with Definitional Vocabulary) had statistically significant correlations in the expected direction with the measures of early academic skills and HTKS scores. The remaining significant correlations were opposite the expected direction, suggesting that once variance due to the general Self-Regulation factor was removed from items on Factors 2–4, the remaining variance represented the degree to which the general factor "over-predicted" item scores on the BRIEF-P. Similar to results from the bifactor model, the results from the structural model using the four factors from the correlated-factors model indicated that only Factor $1_{WM/PO}$ had statistically significant and unique relations with measures of early academic skills and HTKS scores that were in the expected direction. All significant and unique relations between Factors 2–4 and the academic skills and HTKS scores were opposite the hypothesized direction and opposite the direction of the zero-order correlations, indicating that these relations represent suppression effects.

Overall, results of the correlational and structural model analyses with the factors derived from the BRIEF-P in this sample indicated that the measure assesses some component of self-regulation. However, results of these analyses raise questions about the predictive utility and construct validity of some of the subscales of the BRIEF-P. Across analyses, only the general Self-Regulation factor and the factor comprised of items from the BRIEF-P Working Memory and Plan/Organization subscales were uniquely related in expected ways with indices of children's early academic development. Similarly, only the general Self-Regulation factor and the factor comprised of items from the BRIEF-P Working Memory and Plan/Organization subscales were associated with HTKS scores. McClelland et al. (2007) argued that the HTKS assesses multiple aspects of EF, including inhibitory control, working memory, and shifting. Consequently, a failure to find the expected pattern of relations between factors consisting of items primarily from the BRIEF-P Inhibit and Shifting subscales is discrepant from the pattern of findings of previous studies that have found substantial relations between performance-based measures of EF and report-based ratings of behavior self-regulation (e.g., Fuhs et al., 2014; Ponitz et al., 2009). Results of this study suggest that these factors may not validly measure the underlying constructs they are intended to measure. Alternatively, Toplak et al. (2013) argued that performance-based and ratings-based measures of EF measure different levels of EF ability. That is, the HTKS may measure domain-specific aspects of EF in constrained circumstances, whereas the BRIEF-P may measure more general day-to-day EF ability. Similarly, because report-based measures are designed to assess self-regulation through parent and teacher observations of day-to-day activities, these measures capture both self-regulation skills and behaviors within which these skills are demonstrated. It is possible that when the general self-regulation ability is removed from the report-based self-regulation measure, such as in a bifactor model, the variance that remains is attributable to the behavior within which these skills were assessed. This behavior would only be tangentially related to self-regulation, which might explain the lack of significant relations to performance-based measures that are designed to tap only a

specific skill in a situationally constrained manner. A combination of content analysis and additional correlational findings is needed to determine the most likely explanation of these findings.

EF Domains and BRIEF-P Items

As noted above, results of factor-analytic studies with adults and older children indicate that EF is composed of three interrelated skills, inhibitory control (or inhibition), working memory, and shifting (Lehto et al., 2003; Miyake & Friedman, 2012; Miyake et al., 2000). Results of studies of younger children are consistent with a simpler structure of EF, with shifting not emerging as a distinct aspect of EF until early adolescence (Lee, Bull, & Ho, 2013). Studies with preschool populations often report that measures of different aspects of EF are best represented as a single dimension (e.g., Shing, Lindenberger, Diamond, Li, & Davidson, 2010; Wiebe, Espy, & Charak, 2008; Wiebe et al., 2011). However, more recent studies have reported distinct Inhibitory Control and Working Memory factors in children who are at least four years of age (Lerner & Lonigan, 2014; Schoemaker et al., 2012). Although these findings suggest that EF consists of a limited number of skills, all of these structures of EF are based on performance-based EF tasks.

The BRIEF-P purports to measure five aspects of EF: working memory, inhibition, shifting, planning/organization, and emotional control. The first three of these overlap with the components of EF identified in analyses of performance-based EF tasks. Examination of the items that make up the subscales of the BRIEF-P that overlap with these constructs suggests departures between the construct definitions and the BRIEF-P items intended to index them. Several of the items from the Working Memory subscale seem to reflect simple memory (e.g., trouble remembering things, repeats same mistakes, trouble remembering two things) or reflect something only tangentially related to memory (e.g., short attention span, unaware when does well or not well, requires assistance to stay on task). Most items from the Inhibition subscale reflect silly, off task, impulsive behaviors, rather than behaviors that reflect inhibition of a prepotent response to a situation or effectively screening-out competing information. Most of the items from the Shifting subscale appear to reflect shyness or a lack of social flexibility, as these items focus on children's emotional capability to handle new or novel situations, rather than the ability to switch between rules or mental sets. This mismatch between item content and the intended domain may explain why factors made up primarily of items from the Inhibit and Shifting subscales of the BRIEF-P did not relate in expected ways with academic and behavior self-regulation measures.

Limitations and Future Directions

Despite the strengths of this study, including a large sample relative to the number of items on the BRIEF-P, analyses that appropriately accounted for the nested structure of the data, and inclusion of additional measures that allowed an examination of theoretically expected relations between the BRIEF-P and other constructs, the study had a number of limitations. First, although both the correlated-factors model and the bifactor model of the BRIEF-P provided acceptable fits to the data and these models yielded the best possible fits to the data with a limited number of cross-loading items, we did not explore models that may have yielded better fit indices because this would have required eliminating items or dropping

factors (and associated items) from the model. Future studies should be conducted to refine the BRIEF-P to yield a model with improved model fit. Second, because we were interested in comparing optimal structural models of the BRIEF-P instead of refining or defining measurement of a construct, we did not adopt the more conservative strategy of conducting exploratory and confirmatory analyses on different random subsets of the sample. Future studies with additional adequately sized samples should be conducted to refine the item-level structure of the BRIEF-P.

Third, sample demographic data was incomplete, which may have implications for the generalizability of results across ethnicities. However, there is no reason to believe substantial differences between children with and without reported ethnicity information exist, as children were recruited from the same populations, through the same preschools. Furthermore, the sample used in this present study was more ethnically and racially diverse, as compared to the sample used by Isquith et al. (2004). Fourth, we only used teacher-reported BRIEF-P scores in this study. Given findings from previous research that indicate slightly different factor structures for parent and teacher reported BRIEF-P scores (i.e., Duku & Vaillancourt, 2014; Isquith et al.) replication of these findings with parent reported BRIEF-P scores is recommended for future studies. Fifth, we only used one performance-based measure of EF, the HTKS task. Expanding beyond the correlates we selected might help explain some of the more confusing findings, such as the lack of relation between the Inhibition factor and a direct measure of IC. Further, measures that are more similar in item presentation to the BRIEF-P might provide a better context to examine convergent validity. It may be that the performance-based nature of the HTKS renders it difficult to equate to teacher-report measures—although this has not been a problem in previous studies in which at least modest correlations between performance-based and teacher-reported measures of EF have been reported (e.g., Eisenberg et al. 2000). Future studies that include multiple performance- and report-based measures of EF as correlates should be conducted to further assess the convergent validity of the BRIEF-P factors. Finally, the sample size we had available for the HTKS was limited in comparison to the sample size of BRIEF-P. Replication of these findings with a larger and more equivalent sample size is suggested.

Summary and Conclusions

Although the BRIEF-P is a measure commonly used to assess EF in populations of young children and is often used in research to assess the relation between EF skills and other important developmental outcomes (e.g., Clark et al., 2010; Ghassabian et al., 2013), there are few studies that have examined the degree to which it conforms to its hypothesized structure or assessed the predictive utility of its scores with important developmental outcomes. The results of this study yielded mixed findings concerning the structure and predictive utility of the BRIEF-P. Although the BRIEF-P is intended to assess five distinct components of EF, the series of factor analyses conducted in this study indicated that the items of the BRIEF-P do not map onto factors in the way that would be expected based on its purported subscale structure, and the best four-factor solution revealed substantial correlations between factors. Although a bifactor solution identified a general Self-Regulation factor that explained variance in responses across all items, this general factor did not account for all of the overlap among specific factors, resulting in a degree of model

misspecification. Analyses of the correlations of the factors indicated that the majority of the factors had limited convergent validity with academic ability or with a performance measure of behavior self-regulation. Overall, these findings call into question the construct validity of aspects of the measure. Although it appears that, as a whole, the BRIEF-P measures some aspect of the self-regulatory capacity of preschool children and may provide a useful global index of this capacity, it appears unlikely that the specific components of EF that the measure purports to assess are well represented by the items or that the resultant subscales have utility beyond the measurement of a global self-regulatory capacity. Additional work is necessary to refine the measure and to establish the validity of the current or a refined measure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Allan NP, Hume LE, Allan DM, Farrington AL, Lonigan CJ. Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. Developmental Psychology. 2014; 50:2368–2379. [PubMed: 25069051]

Allan NP, Lonigan CJ. Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. Developmental Psychology. 2011; 47:905–915. [PubMed: 21553957]

Allan NP, Lonigan CJ. Exploring dimensionality of effortful control using hot and cool tasks in a sample of preschool children. Journal of Experimental Child Psychology. 2014; 122:33–47. [PubMed: 24518050]

Anderson P. Assessment and development of executive function (EF) during childhood. Child Neuropsychology. 2002; 8:71–82. [PubMed: 12638061]

Best JR, Miller PH, Naglieri JA. Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. Learning and Individual Differences. 2011; 21:327–336. [PubMed: 21845021]

Blair C, Razza RP. Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. Child Development. 2007; 78:647–663. [PubMed: 17381795]

Bodnar EL, Prahme CM, Cutting LE, Denckla MB, Mahone ME. Construct validity of parent ratings of inhibitory control. Child Neuropsychology. 2007; 13:345–362. [PubMed: 17564851]

Bonillo A, Jimenez EA, Ballabriga MC, Capdevila C, Riera R. Validation of Catalan version of BRIEF-P. Child Neuropsychology. 2012; 18:347–355. [PubMed: 21961875]

Bracken, BA. Bracken Basic Concept Scale: Stimulus Manual. Psychological Corporation; 1984.

Bracken BA. Limitations of preschool intstuments and standards for minimal levels of technical adequacy. Journal of Psychoeducational Assessment. 1987; 4:313–326.

Bracken, BA. Bracken school readiness assessment. San Antonio, TX: The Psychological Corporation; 2002.

Brown, T. Confirmatory factor analysis for applied research. New York: Guildford; 2006.

Clark CA, Pritchard VE, Woodward LJ. Preschool executive functioning abilities predict early mathematics achievement. Developmental Psychology. 2010; 46:1176–1191. [PubMed: 20822231]

Diamond A. Development of the ability to use recall to guide action, as indicated by infants' performance on AB. Child Development. 1985; 56:868–883. [PubMed: 4042750]

Diamond, A. The epigenesis of mind: Essays on biology and cognition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1991. Neuropsychological insights into the meaning of object concept development; p. 67-110.

Diamond A, Taylor C. 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. 1996; 29:315–334. [PubMed: 8732806]

Duku E, Vaillancourt T. Validation of the BRIEF-P in a sample of Canadian preschool children. Child Neuropsychology. 2014; 20:358–371. [PubMed: 23721091]

Duncan GJ, Dowsett CJ, Claessen A, Magnuson K, Huston AC, Klebanov P, Pagani LS. School Readiness and Later Achievement. Developmental Psychology. 2007; 43:1428–1446. [PubMed: 18020822]

Eisenberg N, Guthrie IK, Fabes RA, Shepard S, Losoya S, Murphy BC, et al. Prediction of elementary school children's externalizing problem behaviors from attention and behavioral regulation and negative emotionality. Child Development. 2000; 71:1367–1382. [PubMed: 11108101]

Espy K. Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. Developmental Neuropsychology. 2004; 26:379–384. [PubMed: 15276900]

Espy KA, Kaufmann PM, McDiarmid MD, Glisky ML. Executive functioning in preschool children: Performance on A-not-B and other delayed response format tasks. Brain and Cognition. 1999; 41:178–199. [PubMed: 10590818]

Ezpeleta L, Granero R, Penelo E, Osa N, Doménech JM. Behavior Rating Inventory of Executive FUnction-Preschool (BRIEF-P) applied to teachers: Psychometric properties and usefulness for disruptive disorders in 3-year-old preschoolers. Journal of Attention Disorders. 2012:1–3.

Fuhs MW, Farran DC, Nesbitt KT. Prekindergarten children's executive functioning skills and achievement gains: The utility of direct assessments and teacher ratings. Journal of Educational Psychology. 2014

Ghassabian A, Herba CM, Roza SJ, Govaert P, Schenk JJ, Jaddoe VW, Hofman A. Infant brain structures, executive function, and attention deficit/hyperactivity problems at preschool age. A prospective study. Journal of Child Psychology and Psychiatry. 2013; 54:96–104. [PubMed: 22928649]

Gioia, GA., Espy, KA., Isquith, PK. BRIEF-P: Behavior Rating Inventory of Executive Function--preschool Version: Professional Manual. Psychological Assessment Resources; 2003.

Gioia GA, Isquith PK, Guy SC, Kenworthy L. Test review: Behavior Rating Inventory of Executive Function. Child Neuropsychology. 2000; 6:235–238. [PubMed: 11419452]

Gioia GA, Isquith PK, Retzlaff PD, Espy KA. Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. Child Neuropsychology. 2002; 8:249–257. [PubMed: 12759822]

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Converntional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999; 6:1–55.

Isquith PK, Gioia GA, Espy K. Executive function in preschool children: Examination through everyday behavior. Developmental Neuropsychology. 2004; 26:403–422. [PubMed: 15276902]

Jacob R, Parkinson J. The potential for school-based interventions that target executive function to improve academic achievement: A review. Review of Educational Research. 2015:1–41.

Kraybill JH, Bell MA. Infancy predictors of preschool and post-kindergarten executive function. Dev Psychobiol. 2013; 55:530–538. [PubMed: 22692790]

Lonigan, CJ., Wagner, RK., Torgesen, JK., Rashotte, C. Test of Preschool Early Literacy. Austin, TX: ProEd; 2007.

Lee K, Bull R, Ho RMH. Developmental changes in executive functioning. Child Development. 2013; 84:1933–1953. [PubMed: 23550969]

Lehto JE, Juujärvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. British Journal of Developmental Psychology. 2003; 21:59–80.

Lerner MD, Lonigan CJ. Examining dimensionality of executive function among preschool children: Unitary versus distinct abilities. Journal of Psychopathology and Behavioral Assessment. 2014; 36:626–639. [PubMed: 25642020]

MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods. 1996; 1:130–149.

Mahone EM, Hoffman J. Behavior ratings of executive function among preschoolers with ADHD. The Clinical Neuropsychologist. 2007; 21:569–586. [PubMed: 17613979]

McClelland MM, Cameron CE, Connor CM, Farris CL, Jewkes AM, Morrison FJ. Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. Developmental Psychology. 2007; 43:947–959. [PubMed: 17605527]

Matthews JS, Ponitz C, Morrison FJ. Early Gender Differences in Self-Regulation and Academic Achievement. Journal of Educational Psychology. 2009; 101:689–704.

Mcauley T, Chen S, Goos L, Schachar R, Crosbie J. Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? Journal of International Neuropsychological Society. 2010; 16:495–505.

McClelland MM, Cameron CE, Connor CM, Farris CL, Jewkes AM, Morrison FJ. Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. Developmental Psychology. 2007; 43:947–959. [PubMed: 17605527]

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science. 2012; 21:8–14. [PubMed: 22773897]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology. 2000; 41(1):49–100. [PubMed: 10945922]

Muthén, LK., Muthén, BO. Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén; 1998–2011.

Panter JE, Bracken BA. Validity of the Bracken School Readiness Assessment for predicting first grade readiness. Psychology in the Schools. 2009; 46:397–409.

Ponitz CEC, McClelland MM, Jewkes AM, Connor CM, Farris CL, Morrison FJ. Touch your toes! developing a direct measure of behavioral regulation in early childhood. Early Childhood Research Quarterly. 2008; 23:141–158.

Ponitz CC, McClelland MM, Matthews JS, Morrison FJ. A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. Developmental Psychology. 2009; 45:605–619. [PubMed: 19413419]

Schoemaker K, Bunte T, Wiebe SA, Espy KA, Dekovi M, Matthys W. Executive function deficits in preschool children with ADHD and DBD. Journal of Child Psychology and Psychiatry. 2012; 53:111–119. [PubMed: 22022931]

Shaul S, Schwartz M. The role of executive functions in school readiness among preschool-aged children. Reading and Writing. 2013

Sherman EMS, Brooks BL. Behavior Rating Inventory of Executive Function -Preschool Version (BRIEF-P): Test review and clinical guidelines for use. Child Neuropsychology. 2010; 16:503–519.

Shing YL, Lindenberger U, Diamond A, Li S-C, Davidson MC. Memory maintenance and inhibitory control differentiate from early childhood to adolescence. Developmental Neuropsychology. 2010; 35(6):679–697. [PubMed: 21038160]

Sims DM, Lonigan CJ. Inattention, hyperactivity, and emergent literacy: Different facets of inattention relate uniquely to preschoolers' reading related skills. Journal of Clinical Child and Adolescent Psychology. 2013; 42:208–219. [PubMed: 23186142]

Smith-Donald R, Raver CC, Hayes T, Richardson B. Preliminary construct and concurrent validity of the preschool self-regulation assessment (PSRA) for field-based research. Early Childhood Research Quarterly. 2007; 22:173–187.

Steiger JH. Tests for comparing elements of a correlation matrix. Psychological Bulletin. 1980; 87:245–251.

Toplak ME, Bucciarelli SM, Jain U, Tannock R. Executive Functions: Performance-Based Measures and the Behavior Rating Inventory of Executive Function (BRIEF) in Adolescents with Attention Deficit/Hyperactivity Disorder (ADHD). Child Neuropsychology. 2008; 15:53–72.

Toplak ME, West RF, Stanovich KE. Practitioner review: Do performance-based measures and rating of executive function assess the same construct. Journal of Child Psychology and Psychiatry. 2013; 54:131–143. [PubMed: 23057693]

Vitaro F, Brendgen M, Larose S, Tremblay RE. Kindergarten disruptive behaviors, protective factors, and educational achievement by early adulthood. Journal of Educational Psychology. 2005; 97:617–629.

Wanless SB, McClelland MM, Acock AC, Ponitz CC, Son S, Lan X, et al. Measuring behavioral regulation in four societies. Psychological Assessment. 2011; 23:364–378. [PubMed: 21381840]

Wiebe SA, Espy KA, Charak D. Using confirmatory factor analysis to understand executive control in preschool children: I. latent structure. Developmental Psychology. 2008; 44:575–587. [PubMed: 18331145]

Wiebe SA, Sheffield T, Nelson JM, Clark CAC, Chevalier N, Espy KA. The structure of executive function in 3-year-olds. Journal of Experimental Child Psychology. 2011; 108:436–452. [PubMed: 20884004]

Zevenbergen AA, Ryan MM. Gender differences in the relationship between attention problems and expressive language and emerging academic skills in preschool-aged children. Early Child Development and Care. 2010; 180:1337–1348.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Fit statistics for different models of the factor structure of the Behavior Rating Inventory of Executive Function-Preschool Version

Model	Model Type/Description	Model χ^2	df	CFI	TLI	RMSEA	S/WRMR
	Confirmatory						
1.	5-Factor[a]	685.89 ***	34	.89	.98	.09	2.73
2.	1-Factor[a]	1,085.59 ***	21	.82	.94	.14	4.72
	2-Level Exploratory						
3.	1-Factor	44,671.26 ***	3,780	.78	.77	.07	.13
4.	2-Factor	27,969.89 ***	3,718	.87	.86	.05	.09
5.	3-Factor	9,033.02 ***	3,657	.97	.97	.02	.03
6.	4-Factor	7,519.85 ***	3,597	.98	.98	.02	.02
7.	5-Factor	6,767.85 ***	3,538	.98	.98	.02	.02
	Confirmatory						
8.	4-Factor (no crossloads)	834.14 ***	35	.87	.97	.10	2.99
9.	4-Factor (17 crossloads)	605.75 ***	44	.91	.99	.07	2.11
10.	4-Factor Bifactor	492.47 ***	50	.93	.99	.06	1.52

Notes. N = 2,367;

[a]To derive a proper solution, one item had to be excluded from the model because it had a standardized loading greater than 1.0 when included in the model. CFI = comparative fit index; TLI = Tucker-Lewis fit index; RMSEA = root mean square-error of approximation; S/WRMR = Standardized root mean-square residual (SRMR)/weighted root mean-square residual (WRMR).

***
$p < .001$.

Table 2

Factor loadings for Behavior Rating Inventory of Executive Function-Preschool Version items from exploratory two-level factor analysis with four within and one between factors

Item #	Scale	Item Description	F1	F2	F3	F4
14	P/O	Forgets what he/she was supposed to get	**.97***	.02	−.03*	−.10*
37	WM	Forgets in middle of activity	**.95***	−.03*	−.04	−.03
27	WM	Trouble with multi-step activities	**.95***	.05*	−.01	−.10*
2	WM	Trouble remembering two things	**.94***	.02	−.03	−.14*
12	WM	Trouble concentrating during play	**.93***	−.06*	.09*	−.05*
42	WM	Trouble finishing play activities	**.93***	−.07*	.06*	−.02
44	P/O	Cannot find things	**.92***	.07*	−.05*	−.03
49	P/O	Fails to complete tasks	**.91***	−.02	.07*	.00
51	WM	Trouble getting started on activity	**.91***	.06*	.01	−.06*
59	WM	Trouble remembering things	**.89***	.03*	−.21*	.11*
7	WM	Trouble with activities to complete tasks	**.89***	.03*	.15*	−.08*
29	P/O	Trouble solving problems when stuck	**.89***	.08*	−.01	−.09*
17	WM	Repeats same mistakes	**.87***	−.01	.06*	.01
19	P/O	Cannot find personal things	**.86***	.10*	−.07*	−.03
32	WM	Requires assistance to stay on task	**.85***	−.06*	.09*	.10*
55	WM	Cannot finish descriptions of things	**.82***	.05*	−.16*	.10*
9	P/O	Has to be told to start a task	**.79***	.13*	.07*	−.08*
61	WM	Short attention span	**.77***	−.08*	.00	.25*
58	I	Easily sidetracked	**.77***	−.06*	−.02	.27*
47	WM	Cannot maintain topic when talking	**.76***	.01	−.08*	.15*
63	WM	Unaware when performs task right or wrong	**.75***	.10*	−.08*	.18*
4	P/O	Puts things away in a disorganized way	**.75***	−.09*	.27*	.01
57	WM	Unaware when does well or not well	**.74***	.11*	−.07*	.14*
34	P/O	Leaves messes for others to clean up	**.72***	−.08*	.25*	.08*
39	P/O	Gets caught up details of a task	**.63***	.14*	.00	.15*
53	WM	Effort does not match ability	**.56***	.05*	.05*	.16*
38	I	Fails to recognize actions that bother others	**.54***	−.01	**.45***	.14*

Item #	Scale	Item Description	F1	F2	F3	F4
24	P/O	Trouble following established routines	.53*	.17*	.15*	.23*
35	S	Trouble changing activities	.50*	.28*	.20*	.12*
13	I	Needs close supervision	.45*	.08*	.26*	.34*
20	S	Uncomfortable in new situations	-.02*	.97*	-.24*	-.02
10	S	Trouble adjusting to new people	-.04*	.96*	-.17*	-.06*
5	S	Gets upset with new situations	-.05*	.84*	.29*	-.09*
15	S	Gets upset by changes in plans/routines	.02	.81*	.17*	.03
30	S	Disturbed by changes in environment	.09*	.74*	.05	.12*
45	S	Resists change of routine	.15*	.70*	.14*	.08*
40	S	Difficulty joining in at unfamiliar events	.30*	.70*	-.16*	-.12*
25	S	Bothered by intense stimuli	.18*	.59*	.01	.18*
26	EC	Overreacts to small things	.02	.56*	.44*	.17*
36	EC	Reacts more strongly than others	.06*	.55*	.46*	.14*
50	S	Overwhelmed in busy situations	.14*	.52*	-.09*	.46*
46	EC	Stays disappointed for long periods	.01	.51*	.49*	.05
41	EC	Overwhelmed by typical activities	.28*	.50*	.02	.30*
11	EC	Easily upset	-.02	.63*	.70*	-.19*
1	EC	Overreacts to small problems	-.05*	.53*	.66*	-.03
16	EC	Has outbursts with little reason	.04*	.51*	.64*	.01
31	EC	Intense outbursts that end suddenly	.02	.49*	.61*	.02
6	EC	Explosive/angry outbursts	.05*	.40*	.60*	.12*
21	EC	Moody	-.01	.53*	.59*	.05*
33	I	Fails to notice when behavior causes negative reactions	.50*	-.03	.51*	.15*
3	I	Unaware of how behavior affects others	.47*	.02	.48*	.13*
60	I	Becomes too silly	-.01	-.04*	-.03	.95*
18	I	Acts wilder or sillier than others	-.02	.02	.07*	.89*
52	I	Acts wild or out of control	-.01	.03*	.17*	.85*
48	I	Talks/plays loudly	-.06*	.01	.12*	.85*
8	I	Fails to stop laughing when others do	.02	.09*	-.04	.81*
43	I	Gets out of control frequently	.04*	.06*	.20*	.78*

Item #	Scale	Item Description	F1	F2	F3	F4
62	I	Careless/reckless play	.21*	.01	.12*	**.66***
54	I	Has trouble stopping activities	.19*	.01	.26*	**.62***
28	I	Impulsive	.24*	-.01	.26*	**.57***
23	I	Fidgety/restless/squirmy	**.40***	-.06*	.17*	**.49***
22	WM	Makes silly mistakes	**.34***	.04*	.02	**.47***
56	I	Completes tasks/activities too quickly	**.35***	-.02	.00	**.46***

Notes. N = 2,367; Factor loadings equal or greater than .30 shown in bold. P/O = planning and organization; WM = working memory; I = inhibition; S = shifting; EC = emotional control; F1 -F4 = factors 1 – 4.

*
p < .05.

Table 3

Factor loadings for Behavior Rating Inventory of Executive Function-Preschool Version items in four-factor model from confirmatory factor analysis of correlated-factors and bifactor models

BRIEFP-P Item #	BRIEF-P Scale	Correlated-Factors Model				Bifactor Model				
		Factor 1	Factor 2	Factor 3	Factor 4	General Factor	Factor 1	Factor 2	Factor 3	Factor 4
49	P/O	.93				.76	.55			
32	WM	.92				.78	.47			
7	WM	.91				.75	.52			
44	P/O	.91				.69	.62			
58	I	.91				.79	.44			
12	WM	.90				.72	.56			
27	WM	.90				.68	.62			
42	WM	.90				.72	.57			
51	WM	.90				.72	.55			
63	WM	.90				.76	.47			
14	P/O	.89				.67	.62			
17	WM	.89				.76	.48			
24	P/O	.89				.85	.20			
37	WM	.89				.68	.59			
59	WM	.89				.67	.62			
61	WM	.89				.78	.41			
19	P/O	.88				.67	.60			
35	S	.88				.86	.20			
29	P/O	.87				.67	.57			
57	WM	.87				.73	.48			
34	P/O	.86				.77	.36			
9	P/O	.85				.72	.45			
47	WM	.85				.67	.53			
2	WM	.84				.61	.61			
39	P/O	.84				.73	.41			
55	WM	.84				.63	.60			

BRIEFP-P Item #	BRIEF-P Scale	Correlated-Factors Model				Bifactor Model				
		Factor 1	Factor 2	Factor 3	Factor 4	General Factor	Factor 1	Factor 2	Factor 3	Factor 4
4	P/O	.83				.74	.37			
53	WM	.75				.66	.33			
38	I	.72		.48		.93	$.05^{ns}$		−.10	
13	I	.43			.56	.88	.14			.18
45	S		.96			.76		.51		
15	S		.92			.68		.60		
5	S		.89			.62		.66		
30	S		.89			.69		.51		
25	S		.85			.66		.38		
20	S		.82			.45		.80		
10	S		.80			.42		.81		
36	EC		.78	.50		.81		.29	.36	
26	EC		.77	.51		.79		.31	.36	
40	S	$.06^*$	.71	.50		.40	.40	.67		
46	EC		.64			.66		.29	.41	
41	EC		.60			.88		.23		$-.12^{**}$
50	S		.57		.42	.80		.28		$-.03^{ns}$
6	EC		.66	.62		.76		.16	.47	
21	EC		.67	.59		.73		.27	.49	
31	EC		.66	.59		.72		.26	.50	
16	EC		.72	.58		.76		.28	.51	
1	EC		.64	.57		.65		.33	.53	
11	EC		.69	.54		.63		.44	.57	
33	I	.70		.53		.95	$.00^{ns}$		$-.08^*$	
3	I	.68		.48		.89	$.04^{ns}$		$-.02^{ns}$	
43	I				.96	.82				.50
54	I				.96	.86				.34
52	I				.95	.77				.58
18	I				.91	.75				.57

BRIEFP-P Item #	BRIEF-P Scale	Correlated-Factors Model				Bifactor Model				
		Factor 1	Factor 2	Factor 3	Factor 4	General Factor	Factor 1	Factor 2	Factor 3	Factor 4
28	I				.91	.82				.35
62	I				.91	.82				.33
60	I				.87	.70				.58
48	I				.86	.70				.56
8	I				.82	.70				.44
23	I	**.28**			.63	.77	.15			.32
22	WM	**.41**			.40	.68	.24			.24
56	I	**.49**			.29	.64	.29			.19

Notes. Unless marked, all loadings are significant at $p < .001$; loadings in bold are secondary loadings. P/O = Planning and Organization; WM = Working Memory; I = Inhibit; S = Shifting; EC = Emotional Control.

Table 4

Correlations between factors from 4-factor correlated-factors model and bifactor model of Behavior Rating Inventory of Executive Function-Preschool Version (with 17 cross-loaded items) and measures of academic skills and self-regulation

	Correlated-Factors Model				Bifactor Model				
	Factor 1	Factor 2	Factor 3	Factor 4	General Factor	Factor 1	Factor 2	Factor 3	Factor 4
BRIEF-P Factor 1	.98 (.98)								
BRIEF-P Factor 2	.67***	.95 (.95)							
BRIEF-P Factor 3	.20***	.05	.94 (.95)						
BRIEF-P Factor 4	.72***	.48***	.60***	.87 (.94)					
Head-Toes-Knees-Shoulders Task	−.31*** (−.49***)	−.17* (.06)	.07 (.05)	−.11 (.18)	−.20*	−.25***	.00	.11*	.13
TOPEL Definitional Vocabulary	−.36*** (−.57***)	−.24*** (.00)	.05 (−.02)	−.13*** (.29***)	−.24***	−.29***	−.10**	.14***	.15***
TOPEL Print Knowledge	−.28*** (−.56***)	−.10* (.16***)	−.02 (−.08)	−.11*** (.26***)	−.16***	−.26***	.04	.04	.06
TOPEL Phonological Awareness	−.34*** (−.44***)	−.22*** (.00)	−.05 (−.06)	−.19*** (.16*)	−.27***	−.21***	−.04	.07	.08*
Bracken Scales Total Score	−.31*** (−.56***)	−.13*** (.14***)	−.03 (−.07)	−.15*** (.23***)	−.22**	−.23***	.04	.09**	.09*

Notes. N = 2,367; BRIEF-P = *Behavior Rating Inventory of Executive Function-Preschool Version;* HTKS = Head-Toes-Knees-Shoulders task; TOPEL = Test of Preschool Early Literacy. Values along diagonal for correlated-factors model are internal consistency reliabilities (omega values [alpha in parentheses]) for factors. Values in parentheses are standardized coefficients from structural models in which HTKS and academic measures were simultaneously regressed on all factors.

*
p < .05,

**
p < .01,

***
p < .001.