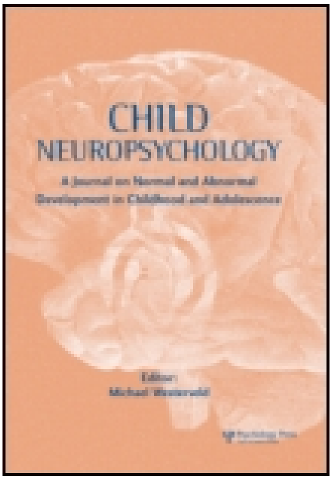


Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncny20

# Factorial validity of the Behavior Rating Inventory of Executive Function (BRIEF)-Teacher form

Christine Peters [a] , James Algina [a] , Stephen W. Smith [a] & Ann P. Daunic [a]
[a] University of Florida , Gainesville , Florida , USA
Published online: 08 Aug 2011.

*Child Neuropsychology*, 2012, 18 (2), 168–181
http://www.psypress.com/childneuropsych
ISSN: 0929-7049 print / 1744-4136 online
http://dx.doi.org/10.1080/09297049.2011.594427

Psychology Press
Taylor & Francis Group



# Factorial validity of the Behavior Rating Inventory of Executive Function (BRIEF)-Teacher form

**Christine Peters, James Algina, Stephen W. Smith, and Ann P. Daunic**

University of Florida, Gainesville, Florida, USA

Deficits in executive function (self-regulatory mechanisms) have been linked with many childhood disorders including attention deficit/hyperactivity disorder (ADHD), autism spectrum disorder, and conduct disorder. Executive functioning is typically assessed by individually administering performance-based measures in a clinical setting. However, performance-based methods are inefficient for school psychologists. A more feasibly implemented measure for applied settings is the *Behavior Rating Inventory of Executive Function* (BRIEF), but researchers have raised questions about the internal validity and the proposed factors. In this study, we examined the factor structure of the teacher form of the BRIEF in a sample of 2,044 general education elementary students and 131 teachers in a multilevel design. Results revealed support for a model with three factors at Level 1 and one general factor at Level 2. The results of our study do not support the current two-factor model of the published BRIEF protocol.

***Keywords:*** BRIEF; Factor analysis; Executive function; Internal validity; Rating scale.

Deficits related to executive function (EF) have been linked to a wide variety of childhood behavior disorders including epilepsy, conduct disorder, autism, Tourette syndrome, and attention deficit/hyperactivity disorder (ADHD; Baltruschat et al., 2011; Barkley, 1997, 2006; Ozonoff & Jensen, 1999; Pennington & Ozonoff, 1996; Sagvolden, Johansen, Aase, & Russell, 2005; Sherman, Slick, & Eyrl, 2006; Sonuga-Barke, 2003). EF has been described as a concept related to self-regulatory behaviors that are necessary to plan and engage in actions that are compliant and goal-directed (Mahone et al., 2002).

Lezak (1995) defined EF as “a collection of interrelated cognitive and behavioral skills that are responsible for purposeful, goal-directed activity and include the highest level of human functioning such as intellect, thought, self-control, and social interaction” (p. 42). Researchers have debated the specific constructs of EF; although most agree that the term “executive function” is an umbrella construct for a collection of interrelated functions (Gioia, Isquith, Retzlaff, & Espy, 2002). Russell Barkley (1997, 2006) and others (e.g., Fuggetta, 2006) are well known for making a connection between EF and ADHD.

Address correspondence to Christine Peters, Yates Elementary School, 301 North Kingsway Road, Brandon, FL 33510, USA. E-mail: drchristinepeters@gmail.com

Barkley's (1997) theory of EF holds that behavioral inhibition is the main underlying factor of EF, and development of behavioral inhibition is necessary for other neuropsychological abilities to function properly. Therefore, ADHD can be viewed as the result of a deficit in behavioral inhibition and, therefore, deficits in EF (Barkley, 2006), and several researchers have demonstrated this connection (Berlin, Bohlin, Nyberg, & Janols, 2004; Fuggetta, 2006; Gooch, Snowling, & Hulme, 2011; Hummer et al., 2011; Jarratt, Riccio, & Siekierski, 2005; Mahone et al., 2002).

Assessment of EF and ADHD is of particular interest to child and school psychologists. EF in children can be assessed using clinical performance-based measures (e.g., Wisconsin Card Sorting Test; Berg, 1948; Heaton, 1981), but they are of limited use in large-scale research and when multiple screenings of children are required. Additionally, the performance-based measures are not always helpful tools for discriminating between disorders or for identifying specific disorders (Barkley & Grodzinsky, 1999; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). When working with children with emotional, behavioral, and/or attentional difficulties, school psychologists typically use omnibus behavior rating scales such as the Behavior Assessment System for Children – Second Edition (BASC-2; Reynolds & Kamphaus, 2004) or the Achenbach System of Empirically Based Assessments (ASEBA; Achenbach & Rescorla, 2001) that include subscales that assess behaviors associated with ADHD and other behavior disorders. They may also assess ADHD symptoms as described in the *Diagnostic and Statistical Manual of Mental Disorders,* fourth edition (*DSM-IV*; American Psychological Association, 1994) more directly by using measures such as the Conners' Teacher Rating Scale, Revised – Short Form (CRS-R; Conners, 2005) or the Conners' ADHD/*DSM-IV* Scales available in parent, teacher, and self-report forms. While all of these measures are widely used and demonstrate good psychometric properties (Achenbach & Rescorla, 2001; Reynolds & Kamphaus, 2004), they are designed to assess symptoms of ADHD and other emotion and behavior disorders rather than children's EF abilities and do not, therefore, directly assess theoretical constructs proposed to explain the development of ADHD (Jarratt et al., 2005).

## The Behavior Rating Inventory of Executive Function (BRIEF)

A promising rating scale that may be used in addition to the aforementioned rating scales is the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000). Gioia and colleagues developed the BRIEF to address concerns about the utility of performance-based EF tests and to assess EF from the everyday, behavioral perspectives of parents and teachers. The authors believe that parents and teachers have unique knowledge about children's behaviors in different environments that are directly relevant to understanding children's EF (Gioia et al., 2000). The BRIEF has scales "similar in concept to, but different in content from, the *DSM-IV* criteria for ADHD" (Mahone et al., 2002, p. 646). Anderson, Anderson, Northam, Jacobs, and Mikiewicz (2002) state that the BRIEF is the first instrument designed to assess behavioral aspects of EF in children.

The BRIEF consists of a parent form and teacher form, both of which contain 86 items rated on a 3-point Likert scale (1 = never, 2 = sometimes, 3 = often). The BRIEF produces scores for eight clinical scales that measure different aspects of EF: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. As shown in Figure 1, the clinical scales form two broad indexes, Behavioral Regulation and Metacognition. The Behavioral Regulation Index (BRI) consists of the

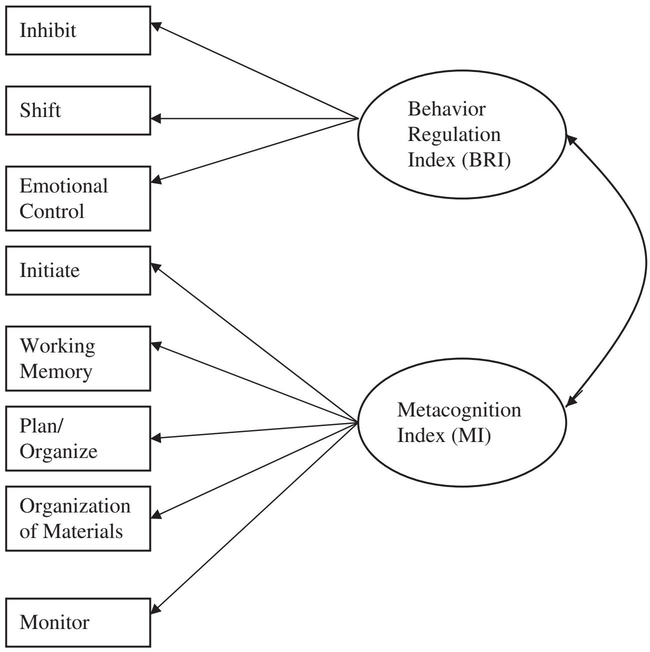


**Figure 1** Oblique two-factor model.



Inhibit, Shift, and Emotional Control subscales that are all related to a child's ability to use inhibitory control to shift cognitive set and to manage emotions and behavior. The Metacognition Index (MI) consists of the Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor subscales. The MI reflects a child's ability to cognitively self-manage tasks and to monitor his or her performance (Gioia et al., 2000). These indexes can be combined to calculate a general overall score, the Global Executive Composite.

The normative sample for the BRIEF Teacher Form included children ages 5 through 18. Seven hundred and twenty teachers of 403 girls and 317 boys completed the teacher form, and the data in the normative sample were weighted to reflect a representation of students from racial/ethnic backgrounds and socioeconomic status that matched the U.S. population (Gioia et al., 2000). The accompanying parent form was normed using parent ratings of 1419 control children and 852 children from clinical groups. Data from the normative sample were used to conduct principal factor analysis (PFA) as an exploratory method. Results for both the parent and teacher forms indicated good fit for an oblique two-factor model. The two oblique factors identified were the BRI and the MI, and they were moderately correlated ($r = .67$ and $r = .62$ on the parent and teacher forms, respectively; Gioia et al., 2000).

Research on the utility and validity of the BRIEF has mostly revealed positive results. Jarratt et al. (2005) found support for concurrent validity comparing scores on the BASC (Reynolds & Kamphaus, 1992) and the BRIEF. Support has also been found comparing scores on the CRS-R (Conners, 2005) and the BRIEF (Alloway et al., 2009). Anderson et al. (2002) used the BRIEF and traditional performance-based measures of EF with children and found that the BRIEF was efficient in detecting behavioral symptoms consistent with deficits in EF. However, Anderson and colleagues noted that the relationships between scores on the BRIEF subscales and scores from performance-based measures were inconsistent, suggesting that the BRIEF is tapping separate functions. Similar studies have found inconsistent correspondence between the BRIEF scores and results on performance-based measures (Alloway et al., 2009; Mahone et al., 2002; Toplak, Bucciarelli, Jain, & Tannock, 2009).

One specific component of validity is the notion of internal consistency that is often measured via factor analytic techniques. Two studies have examined the internal factor structure of the BRIEF and results revealed support for the individual scales but questionable support for the indexes. Gioia, Isquith, Retzlaff, and Espy (2002) conducted a confirmatory factor analysis (CFA) on the BRIEF Parent Form with children from a clinical sample and found that the proposed oblique two-factor model was a poor fit to the data. Gioia and Isquith (2002) had previously suggested that the Monitor scale may actually measure two distinct dimensions and subsequently divided it into two scales, one being Self-Monitor (monitoring of personal behavioral activities) and the other Task-Monitor (monitoring of task-related activities).

In their CFA study, Gioia et al. (2002) tested four models and identified the best fitting model as one comprising three factors: the Inhibit and Self-Monitor scales, indicating a latent Behavioral Regulation factor; the Emotional Control and Shift scales, indicating a latent Emotional Regulation factor; and the Initiate, Working Memory, Plan/Organize, Organization of Materials, and Task-Monitor scales, indicating a latent Metacognition factor. In a post hoc analysis, the authors significantly improved the fit of the three-factor model by allowing the residuals for the Inhibit scale to correlate with residuals from the Working Memory, Organization of Materials, and Emotional Control scales based on theory and statistics (see Gioia et al., 2002, for a description). To address the difference in model fit from previous analyses of the BRIEF, the authors recommended (a) interpreting the standard published eight scales and two factors (BRI, MI) or (b), considering the findings from their study, separating the Monitor scale and interpreting the three-factor model.

Slick, Lautzenhiser, Sherman, and Eyrl (2006) conducted a PFA with results from parent ratings of 80 children with epilepsy on the BRIEF. The authors used a forced two-factor extraction with oblique rotation and found that four scales loaded strongly on the first factor (Plan/Organize, Working Memory, Initiate, Organization of Materials), three scales loaded strongly on the second factor (Emotional Control, Shift, Inhibit), and the Monitor scale loaded strongly on both factors (Slick et al., 2006). The finding that the Monitor scale loaded on both factors is consistent with research by Gioia and Isquith (2002) and Gioia et al. (2002).

As discussed, the published factor analyses on the BRIEF conducted by Gioia et al. (2002) and Slick et al. (2006) used the parent form of the BRIEF with children from clinical samples. To date, it does not appear that researchers have conducted an examination of the factor structure of the BRIEF teacher form on data collected from a nonclinical sample or by using multilevel factor analysis. Given the conflicting evidence about the factor structure of the BRIEF, and with little knowledge of internal validity for general education students, the purpose of our study was to examine the evidence for validity based on factor structure of the BRIEF for use with general education children in the school setting. In addition, given the nature of our data, we were able to examine the factor structure at the child level (Level 1) and the classroom level (Level 2) to determine whether the two levels contain similar factors. Specifically, we used both the oblique two-factor model structure (Gioia et al., 2000) and the oblique three-factor model (Gioia et al., 2002) with data from the BRIEF Teacher Form.

## METHODS

We collected the data for this study as part of a larger randomized control field trial of the effectiveness of a 26-lesson social problem-solving curriculum (*Tools for Getting Along*) taught in fourth and fifth grade classrooms on chronic classroom disruption and aggression (cf. Daunic, Smith, Brank, & Penfield, 2006; Daunic et al., 2007). *Tools for Getting Along* is a cognitive-behavioral intervention aimed at teaching children to use a series of social problem-solving steps in anger-provoking situations. As part of the field trial, we recruited schools based on lunch status (i.e., percentage of free and reduced lunch), targeting schools with higher percentages of low-income families. During the 3-year study, 18 elementary schools in six school districts agreed to participate. The sample consisted of 131 fourth and fifth grade classrooms where each teacher completed several assessments on their students pre- and postintervention. We analyzed only pretest BRIEF data for students with consent, because teachers who delivered *Tools for Getting Along* might have interpreted items differently at postassessment than teachers in the control condition, resulting in different factors.

### Participants and Instruments

Participants in the 3-year study were 2,101 fourth and fifth grade students from suburban and rural areas in a large state in the southeast. Specifically, the students were 51.5% Caucasian, 27.7% African American, 8.7% Hispanic, 3.1% mixed, 0.4% Indian, 0.3% Asian, and 8.2% missing race/ethnicity data. Additionally, the sample was 46.4% ($n = 975$) female and 45.4% ($n = 954$) male (with 8.2% missing gender data), and 72.9% of participants were eligible to receive free or reduced lunch. Data on students' potential special education eligibility were not available at the time of the current study; although some of the schools operated under an inclusion model so some of the students in the general education classrooms received special education services. Teachers completed the pretest BRIEF Teacher Form in November of the academic year, well past the minimum 1 month of daily contact recommended in the BRIEF manual. Given the nature of the primary study, teachers were notified whether they were in the treatment or control group prior to completing the BRIEF. After eliminating cases with entirely missing data for the BRIEF, we conducted factor analyses with a sample of 2,044 students and 131 teachers.

### Statistical Analyses

In the present study, the data have a multilevel structure: Children are nested in classrooms and each teacher completes assessments on students from one classroom only. Conducting a single-level CFA would entail the assumption that clinical scale scores for different students are statistically independent, an assumption that would be violated if mean scale scores vary reliably across teachers. Alternatively, a two-level CFA, which posits a within-teacher factor model and a between-teacher factor model, properly addresses a potential lack of independence. The advantages of a two-level CFA are fourfold. First, and perhaps most important, a two-level CFA allows for the possibility of a different number of factors and/or different estimates of factor loadings in the within- and between-teacher models. For example, it is possible that Gioia's three-factor model could fit well as a within-teacher model but not as a between-teacher model. As pointed out by Muthen (1991, 1994), it is not uncommon in multilevel factor analysis that support will be found for a complex Level 1 (within-teacher) model with several factors and a simpler Level 2 (between-teacher) model with fewer factors. Second, due to the incorporation of the within- and between-teacher models, a multilevel-CFA can fit the data better than a single-level CFA. Third, if different within- and between-teacher models are required, the factors in a single-level CFA will be a blend of the within- and between-teacher factors and may be less interpretable than the factors from the multilevel CFA. Fourth, estimation of standard errors of parameter estimates, chi-square statistics, and goodness-of-fit indexes can be improved by using a multilevel CFA.

Our research design actually represents a three-level model, as students are nested within classrooms, and classrooms are nested within schools. We ignored the three-level nature of the data, in part because there was not a sufficient number of schools to conduct a three-level CFA and in part because a three-level CFA is not currently implemented in commercially available software.

We obtained scale scores by summing the raw scores from questions assigned to each scale. As suggested by the BRIEF manual (Gioia et al., 2000), missing items were assigned a raw score of one, and scores were not calculated for scales that contained more than two missing items. Overall, 10 scales were used: the original eight scales and, in some models, the new Self-Monitor and Task-Monitor scales were used in place of the Monitor scale (Gioia & Isquith, 2002). The latent factor structure of eight or nine BRIEF scales (depending on the model) was examined via a maximum likelihood CFA using the Mplus version 5.2 statistical software program (Muthen & Muthen, 1998–2004) and a full information robust maximum likelihood estimation. This method includes data from students who have scores on all clinical scales and from students who have incomplete data. The robust procedure adjusts standard errors and goodness-of-fit chi-square statistics for non-normality.

We examined four models representing different factor structures of the BRIEF. In Models 1 and 2, the within-teacher model comprised the theoretical factor structure of the published BRIEF with two oblique factors, where three scales (Inhibit, Shift, Emotional Control) loaded on the latent Behavioral Regulation factor, and five scales (Initiate, Working Memory, Plan/Organize, Organization of Materials, Monitor) loaded on the latent Metacognition factor. Figure 1 depicts the two-factor structure of the within-teacher model (i.e., the published protocol). In Model 1 the between-teacher model posited a general factor with all eight scales loading on one factor, and, in Model 2, the between-teacher model consisted of the two-factor structure of the published protocol. In Models 3 and 4, the within-teacher model was the three-factor model from Gioia et al. (2002) where two scales (Inhibit, Self-Monitor) loaded on the latent Behavioral Regulation factor, two scales (Shift, Emotion Regulation) loaded on the latent Emotion Regulation factor, and five scales (Initiate, Working Memory, Plan/Organize, Organization of Materials, Task-Monitor) loaded on the latent Metacognition factor. Figure 2 depicts the within-teacher model for Models 3 and 4. In Model 3, the between-teacher model was a single-factor model in which all nine scales loaded on one factor, whereas the between-teacher model for Model 4 consisted of the three-factor model. Models with correlated errors (as published in Gioia et al., 2002) were attempted but are not reported because the models produced warnings and estimation did not converge properly.

We used the chi-square and degrees of freedom goodness-of-fit test to determine goodness of fit for the eight models. In addition, we used the following goodness-of-fit indexes to compare models: Bentler's comparative fit index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized

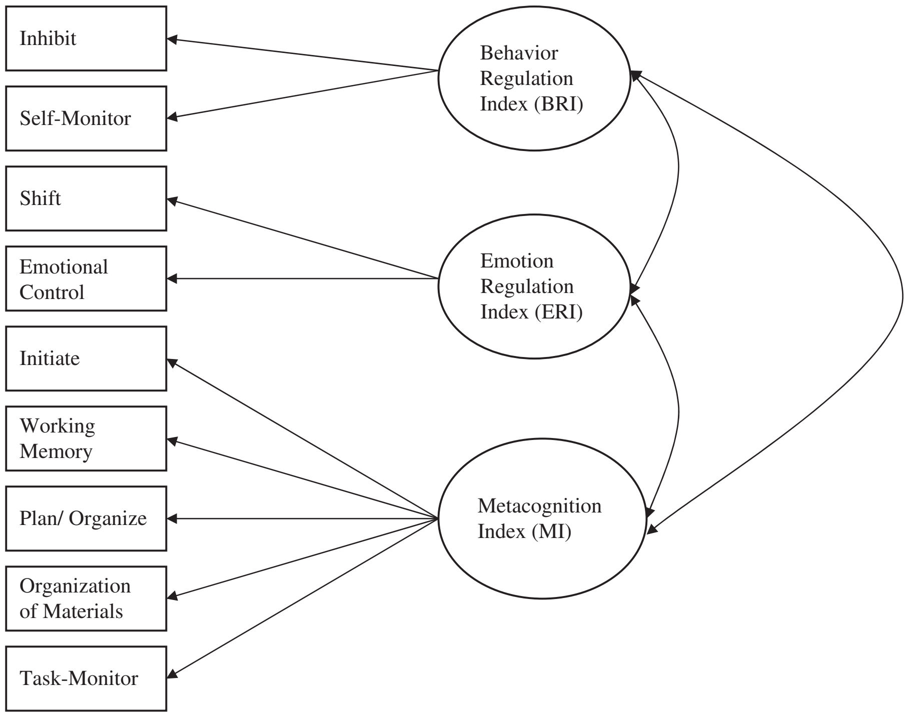


**Figure 2** Oblique three-factor model.

Root Mean Square Residual (SRMR). Criteria for good fit include CFI and TLI $\geq$ .95, RMSEA $\leq$ .06, and SRMR $\leq$ .09 (Hu & Bentler, 1999).

## RESULTS

Table 1 shows the means, within-teacher standard deviations, between-teacher standard deviations, intraclass correlation coefficients, and minimum and maximum for BRIEF scale scores. The within-teacher standard deviations measure the extent of individual differences within classrooms, while the between-teacher standard deviations measure the extent of differences between classrooms. The descriptive statistics for the BRIEF scales suggest that sample mean scores on all scales were much lower than the mean scores reported for clinical samples (Gioia et al., 2000, 2002; Slick et al., 2006), and sample means were slightly higher than the mean scores reported for the normative sample (for children 9 to 13 years, mean scores ranged from 7.89 to 12.74). This finding may be due to the fact that we recruited schools with a high percentage of students on free or reduced price lunch and that some of the schools practiced inclusion of special education students in general education classrooms (all students with informed consents were included in the study). Inspection of the within- and between-teacher standard deviations indicates there is more variability within teachers than between teachers. Nevertheless, the intraclass correlation coefficients range from .167 for the Inhibit scale to .323 for the Shift scale, indicating that 17% to 32% of the variance in scale scores is at the teacher level. This finding highlights the need to conduct a multilevel factor analysis with these nested data.

**Table 1** BRIEF Scale Means and Standard Deviations.

| BRIEF Scale | Mean | *SD* Within | *SD* Between | Intraclass Correlation | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Inhibit | 15.149 | 5.608 | 2.51 | 0.167 | 10.00 | 30.00 |
| Shift | 13.945 | 4.037 | 2.789 | 0.323 | 10.00 | 30.00 |
| Emotional Control | 12.801 | 4.580 | 2.230 | 0.192 | 9.00 | 27.00 |
| Initiate | 11.335 | 3.535 | 1.883 | 0.221 | 7.00 | 21.00 |
| Working Memory | 15.412 | 4.906 | 2.512 | 0.208 | 10.00 | 30.00 |
| Plan/Organize | 15.590 | 4.730 | 2.648 | 0.239 | 10.00 | 30.00 |
| Organization of Materials | 10.141 | 3.506 | 1.661 | 0.183 | 7.00 | 21.00 |
| Monitor | 15.412 | 4.880 | 2.610 | 0.222 | 10.00 | 30.00 |
| Self-Monitor | 9.255 | 3.192 | 1.660 | 0.213 | 6.00 | 18.00 |
| Task Monitor | 6.739 | 2.128 | 1.064 | 0.200 | 4.00 | 12.00 |

**Table 2** Correlations of BRIEF Observed Variables for Within-Teacher (Level 1).

| BRIEF Scale | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Inhibit | 1.000 | | | | | | | | | |
| 2. Shift | .712 | 1.000 | | | | | | | | |
| 3. Emotional Control | .787 | .849 | 1.000 | | | | | | | |
| 4. Initiate | .667 | .675 | .585 | 1.000 | | | | | | |
| 5. Working Memory | .683 | .685 | .576 | .883 | 1.000 | | | | | |
| 6. Plan/ Organize | .682 | .696 | .589 | .897 | .898 | 1.000 | | | | |
| 7. Organization of Materials | .650 | .595 | .532 | .695 | .767 | .793 | 1.000 | | | |
| 8. Monitor | .869 | .734 | .731 | .827 | .823 | .844 | .750 | 1.000 | | |
| 9. Self-Monitor | .900 | .730 | .771 | .714 | .707 | .713 | .641 | – | 1.000 | |
| 10. Task-Monitor | .641 | .588 | .520 | .826 | .824 | .865 | .757 | – | .669 | 1.00 0 |

Tables 2 and 3 display the within-teacher and between-teacher correlation matrices for the 10 observed variables. As shown, more variability among correlations between scales is evident within level (Table 2) than between level (Table 3), and the between-teacher correlation coefficients tend to be larger than the within-teacher correlation coefficients. These results suggest that the factor structure will be different within teachers versus between teachers.

Table 4 summarizes the results from the multilevel CFA on the four models. Model 1, with two factors at Level 1 (within-teachers) and one factor at Level 2 (between-teachers), yielded a $\chi^2(39) = 1449.547$, indicating poor fit. In addition, CFI = 0.880, TLI = 0.827, and RMSEA = 0.133, also indicating poor fit. SRMR for the within and between models was adequate (SRMR within = 0.066; SRMR between = 0.039). Results for Model 2, with two factors at both levels, as proposed by the published protocol, also indicated poor fit. For this model, $\chi^2(38) = 1417.818$, CFI = 0.882, TLI = 0.827, and RMSEA = 0.133. Again, SRMR was adequate (SRMR within = 0.065; SRMR between = 0.034). Results for Model 3, with three factors at Level 1 (Gioia et al., 2002) and one general factor at Level 2, were promising. Model 3 yielded a $\chi^2(51) = 691.787$, indicating poor fit. However, CFI = 0.954, TLI = 0.934, SRMR within = 0.034, and SRMR between = 0.036, indicating adequate fit. RMSEA for Model 3 remained high; although it was close to the criterion of 0.06 (RMSEA = 0.078). Model 4 consisted of three factors at both Levels 1 and 2. The results for Model 4 were also promising: CFI = 0.957, TLI = 0.935, SRMR within

**Table 3** Correlations of BRIEF Observed Variables for Between-Teacher (Level 2).

| BRIEF Scale | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Inhibit | 1.000 | | | | | | | | | |
| 2. Shift | .893 | 1.000 | | | | | | | | |
| 3. Emotional Control | .900 | .931 | 1.000 | | | | | | | |
| 4. Initiate | .881 | .853 | .765 | 1.000 | | | | | | |
| 5. Working Memory | .887 | .904 | .804 | .942 | 1.000 | | | | | |
| 6. Plan/ Organize | .865 | .903 | .795 | .953 | .950 | 1.000 | | | | |
| 7. Organization of Materials | .831 | .904 | .793 | .845 | .892 | .911 | 1.000 | | | |
| 8. Monitor | .938 | .880 | .837 | .947 | .910 | .926 | .855 | 1.000 | | |
| 9. Self-Monitor | .951 | .893 | .863 | .883 | .881 | .868 | .817 | – | 1.000 | |
| 10. Task-Monitor | .812 | .764 | .702 | .937 | .855 | .912 | .818 | – | .819 | 1.000 |

**Table 4** Summary of Fit Statistics for the Four BRIEF Models.

| Model | $\chi^2$ | *df* | CFI | TLI | RMSEA | SRMR[a] |
|---|---|---|---|---|---|---|
| 1. Level 1: Two factors | 1449.547* | 39 | 0.880 | 0.827 | 0.133 | W: 0.066 |
| Level 2: One factor | | | | | | B: 0.039 |
| 2. Level 1: Two factors | 1417.819* | 38 | 0.882 | 0.827 | 0.133 | W: 0.065 |
| Level 2: Two factors | | | | | | B: 0.034 |
| 3. Level 1: Three factors | | 51 | 0.954 | 0.934 | 0.078 | W: 0.036 |
| Level 2: One factor | 691.787* | | | | | B: 0.055 |
| 4. Level 1: Three factors | | 48 | 0.957 | 0.935 | 0.078 | W: 0.034 |
| Level 2: Three factors | 647.106* | | | | | B: 0.030 |

*Note.* $N = 614$ for all models. Mplus produced warnings about this model as being nonidentified with a negative residual variance for the shift scale.

[a]W: SRMR for within; B: SRMR for between.

$^*p \leq .001$.

= 0.034, and SRMR between = 0.030. Model 4 also yielded a $\chi^2(48) = 647.106$ and RMSEA = 0.078, indicating cautions about the model. Additionally, the results for the between-teacher model contained a negative residual variance for the Shift scale, suggesting that the three-factor between-teacher model contained too many factors.

Finally, it is worth noting that we analyzed these data in a single-level, three-factor model. Information criteria indicated better fit for the multilevel model. For example, Akaike's information criterion is a measure of fit in which smaller values indicate better fit, and additional parameters are likely to produce larger values (i.e., penalizes for estimation of multiple parameters). Akaike's information criterion was 83016.540 for the single-level, three-factor model and 81554.675 for the multilevel model with three within-teacher factors and one between-teacher factor, indicating that the multilevel CFA fit better despite the penalty for the extra parameters in the multilevel model.

Table 5 displays standardized factor loadings for the multilevel model with three within-teacher factors and one between-teacher factor. In the within-teacher model, all scales have high and statistically significant loadings on their intended factor. In the between-teacher model, all scales have high loadings on the single factor. The within-teacher correlations for the factors are reported in Table 6. All three correlation coefficients are substantial, with the strongest relationship between the two self-regulation types.

**Table 5** Standardized Factor Loadings for Multilevel Model with Three Within-Teacher and One Between-Teacher Factors.

| | Within-Teacher | | | Between-Teacher |
|---|---|---|---|---|
| Scale | Behavior Regulation Index | Emotion Regulation Index | Metacognition Index | General |
| Inhibit | .94* | | | .97* |
| Self-Monitor | .96* | | | .97* |
| Shift | | .90* | | .93* |
| Emotional Control | | .94* | | .97* |
| Initiate | | | .93* | .97* |
| Working Memory | | | .94* | .98* |
| Plan/Organize | | | .96* | .92* |
| Organize Materials | | | .81* | .91* |
| Task-Monitor | | | .89* | .95* |

$^{*}p < .05$.

**Table 6** Correlation Coefficients for the Within-Teacher Factors.

| | Behavior Regulation Index | Emotion Regulation Index | Metacognition Index |
|---|---|---|---|
| Behavior Regulation Index | 1.00 | – | |
| Emotion Regulation Index | .86* | 1.00 | – |
| Metacognition Index | .77* | .69* | 1.00 |

$^{*}p < .05$.



## DISCUSSION

The BRIEF is a promising behavior rating scale for the ecological assessment of EF, especially of inhibition and working memory in children thought to have EF deficits (Gioia et al., 2000). Researchers have shown that the BRIEF can identify students with behavioral difficulties and deficits in EF (Alloway et al., 2009; Anderson et al., 2002; Jarratt et al., 2005; Mahone et al., 2002; Toplak et al., 2009) showing promise as an additional assessment tool for the identification of ADHD in school-aged children. Research on the internal factor structure and construct validity of the BRIEF, however, has raised questions. Gioia et al. (2002) and Slick et al. (2006) failed to find support for the two-factor model originally proposed by the authors and represented in the published protocol. Gioia et al. (2002) found support for a three-factor model comprising nine scales as opposed to eight, with the Monitor scale split into Self-Monitor and Task-Monitor scales. The Gioia and Slick research groups used the BRIEF Parent Form on clinical samples. The purpose of our study was to replicate and expand on their research on the factor structure of the BRIEF using the teacher form and a nonclinical, general education sample. In addition, we examined the multilevel nature of the data by investigating different models at Level 1, the within-teacher model, and Level 2, the between-teacher model.

Our results for the within-teacher model with two factors replicated the findings that the two-factor model of the published protocol did not fit the data well. This result occurred with both a single between-teacher factor and two between-teacher factors. We found that a three-factor, within-teacher model with two scales loading on the latent Behavioral

Regulation factor, two scales loading on the latent Emotion Regulation factor, and five scales loading on the latent Metacognition factor fit better than the two-factor model and fit reasonably well, according to criteria suggested by Hu and Bentler (1999). The improved fit occurred when the three-factor, within-teacher model was paired with a one-factor, between-teacher (Model 3) and a three-factor between-teacher model (Model 4). Both Models 3 and 4 produced adequate fit indices. Model 4, however, resulted in a negative residual variance for one of the scales, indicating a preference for Model 3.

The intraclass correlation coefficients for the scales indicated reliable between-teacher differences in mean scale scores. The between-teacher factor model suggested these between-teacher differences can be accounted for by a single factor. The extent to which the between-teacher differences reflect real differences among the students in the various classrooms or differences in the rating behavior of teachers is unknown and is an important direction for future research on the BRIEF and other rating scales used for psychological assessment in schools.

The results for the between-teacher models suggest that classroom average scale scores define one factor. This has implications for interpreting school-based studies using scores on the BRIEF. For example, consider a study conducted on teacher practices (e.g., classroom discipline) that may predict outcomes related to student behavioral and emotional self-regulation. Because the relationship of a measure of teacher practice to student-level measures must be due to the relationship of teacher practice to classroom means of the student-level measures, results using the BRIEF would likely show that teacher practice has a similar relationship to behavioral and emotional self-regulation. Whether this finding would be an artifact of using the BRIEF or a correct assessment of the relationship of teacher practice to behavioral and emotional self-regulation would be an open question. It should be noted that concern about the general factor in the between-teacher model is only relevant when teachers complete the BRIEF, as the multilevel nature of the data is not relevant to the parent form. Finally, we found large correlations between within-teacher factors (i.e., behavior regulation, emotion regulation, metacognition) with the highest between behavior regulation and emotion regulation. These high interfactor correlations are consistent with the findings of current EF researchers who have described the interrelated nature of emotional and behavioral self-regulation and more strictly cognitive processes (see, e.g., Blair & Diamond, 2008; Rueda, Posner, & Rothbart, 2005; Zelazo & Cunningham, 2007).

Of slight concern for researchers and practitioners currently using the BRIEF is that results from two studies in addition to the current study have not fully supported the internal factor structure of the current published protocol. More specifically, while the clinical scales appear to have internal consistency, there has been more support for three factors than two factors. Although Gioia et al. (2002) have recommended interpreting the eight clinical scales and two composite scales as published, our data do not fully support the authors' claim regarding the composite scales. Instead, we recommend that future revisions for the BRIEF consider a three-factor model. In the meantime, practitioners may continue to use the computer scoring program that provides scores and excellent descriptions for each of the eight subscales (the description of the monitor subscale discusses the distinction between self- and task-monitoring as well). However, we urge practitioners to be cautious when using information about the BRI and MI and present information to parents and teachers with emphasis on the individual scales rather than the indexes. In addition, we suggest that more research be conducted on the theory underlying the BRIEF and/or possible revisions be made to the protocols and scoring program.

Without revisions to the BRIEF, a concern regarding calculating raw scores for new scales and recalculating indexes relates to a current lack of normative data for a three-factor model, without which conclusions from test results are inappropriate (American Educational Research Association; American Psychological Association; National Council on Measurement in Education, 2002). It would be beneficial for researchers to use existing normative data to construct the required norms or to collect new normative data on both the parent and teacher forms and to conduct further validation studies using the three-factor model to test whether use of the revised instrument is in line with the *Standards for Educational and Psychological Testing.* Although the three-factor model represents minor changes to the subscales, the model proposes a third latent factor of EF. Additional research on the third factor, emotional regulation, is needed before clinical and practical interpretations are warranted.

This study has several strengths including the large sample size and use of multilevel confirmatory and exploratory factor analytic techniques; however, it is not without limitations. While the BRIEF is normed for use with students ages 5–18, our study was focused on students in fourth and fifth grades (approximately 9–12 years old) limiting generalization to younger and older populations. The generalization of our findings is also limited because our sample was restricted to a large state in the southeast. As such, we suggest additional studies that include less restricted age ranges and geographic regions across the United States. Finally, a large proportion of students in this study were eligible for free or reduced-priced lunch, indicating a low level of family income. Thus, the generalization of our findings may be limited to students from low income and future studies are encouraged to use a more diverse sample with regards to family income.

Original manuscript received January 17, 2011
Revised manuscript accepted May 15, 2011
First published online August 5, 2011

## REFERENCES

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

Alloway, T. P., Gathercole, S. E., Holmes, J., Place, M., Elliott, J. G., & Hilton, K. (2009). The diagnostic utility of behavioral checklists in identifying children with ADHD and children with working memory deficits. *Child Psychiatry and Human Development*, *40*, 353–366.

American Educational Research Association; American Psychological Association; National Council on Measurement in Education. (2002). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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.

Baltruschat, L., Hasselhorn, M., Tarbox, J., Dixon, D. R., Najdowski, A. C., Mullins, R. O., et al. (2011). Assessing working memory in children with autism through behavioral intervention. *Research in Autism Spectrum Disorders*, *5*, 267–276.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65–94.

Barkley, R. A. (2006). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (3rd ed.). New York, NY: Guilford Press.

Barkley, R. A., & Grodzinsky, G. M. (1999). Predictive power of frontal lobe tests in the diagnosis of attention-deficit hyperactivity disorder. *Clinical Neuropsychologist*, *13*, 12–21.

Berg, E. A. (1948). A simple objective treatment for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15–22.

Berlin, L., Bohlin, G., Nyberg, L., & Janols, L. O. (2004). How well do measures of inhibition and other executive function discriminate between children with ADHD and controls? *Child Neuropsychology*, *10*, 1–13.

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, *20*, 899–911.

Conners, K. (2005). *Conners teacher rating scale-revised-short*. New York, NY: Multi-Health Systems.

Daunic, A. P., Smith, S. W., Brank, E. M., & Penfield, R. D. (2006). Classroom based cognitive-behavioral intervention to prevent aggression: Efficacy and social validity. *Journal of School Psychology*, *44*, 123–139.

Daunic, A. P., Smith, S. W., Garvin, C., Naranjo, A., Barber, B., Becker, M., et al. (2007, July). *First year progress of a universal cognitive-behavioral intervention for elementary students to reduce disruptive/aggressive behavior.*(Poster Presentation). U.S. Department of Education, Institute of Education Sciences 2007 Research Conference, Washington, DC.

Fuggetta, G. P. (2006). Impairment of executive functions in boys with attention deficit/hyperactivity disorder. *Child Neuropsychology*, *12*, 1–21.

Gioia, G. A., & Isquith, P. K. (2002). Two faces of Monitor: Thy self and thy task [Abstract]. *Journal of the International Neuropsychological Society*, *8*, 229.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function Professional Manual*. Lutz, FL: Psychological Assessment Resources.

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

Gooch, D., Snowling, M., & Hulme, C. (2011). Time perception, phonological skills and executive function in children with dyslexia and/or ADHD symptoms. *Journal of Child Psychology and Psychiatry*, *52*, 195–203.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test (WCST)*. Odessa, FL: Psychological Assessment Resources.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Hummer, T. A., Kronenberger, W. G., Wang, Y., Dunn, D. W., Mosier, K. M., Kalnin, A. J., et al. (2011). Executive functioning characteristics associated with ADHD comorbidity in adolescents with disruptive behavior disorders. *Journal of Abnormal Child Psychology*, *39*, 11–19.

Jarratt, K. P., Riccio, C. A., & Siekierski, B. M. (2005). Assessment of Attention Deficit Hyperactivity Disorder (ADHD) using the BASC and BRIEF. *Applied Neuropsychology*, *12*, 83–93.

Lezak, M. (1995). *Neuropsychological assessment*. New York, NY: Oxford University Press.

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., et al. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology*, *17*, 643–662.

Muthen, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338–354.

Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*, 376–398.

Muthen, L. K., & Muthen, B. O. (1998–2004). Mplus version 5.0 [Computer software]. Los Angeles, CA: Muthen & Muthen.

Ozonoff, S., & Jensen, J. (1999). Specific EF profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, *29*, 171–177.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *37*, 51–87.

Reynolds, C. R., & Kamphaus, R. W. (1992). *The Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Services.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: American Guidance Services.

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology*, *28*, 573–594.

Sagvolden, T., Johansen, E. B., Aase, H., & Russell, V. A. (2005). A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *Behavioral and Brain Sciences*, *28*, 397–419.

Sherman, E. M. S., Slick, D. J., & Eyrl, K. M. (2006). Executive dysfunction is a significant predictor of poor quality of life in children with epilepsy. *Epilepsia*, *47*, 1936–1942.

Slick, D. J., Lautzenhiser, A., Sherman, E. M. S., & Eyrl, K. (2006). Frequency of scale elevations and factor structure of the Behavior Rating Inventory of Executive Function (BRIEF) in children and adolescents with intractable epilepsy. *Child Neuropsychology*, *12*, 181–189.

Sonuga-Barke, E. J. S. (2003). The dual pathway model of AD/HD: An elaboration of neurodevelopmental characteristics. *Neuroscience and Biobehavioral Reviews*, *27*, 593–604.

Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2009). Executive functions: Performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology*, *15*, 53–72.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*, 1336–1346.

Zelazo, P. D., & Cunningham, W. A. (2007). Executive function: Mechanisms underlying emotion regulation. In J. Gross (Ed.), *Handbook of emotion regulation* (pp. 135–158). New York, NY: Guilford Press.