Published in final edited form as:

J Neuropsychol. 2017 June ; 11(2): 252–276. doi:10.1111/jnp.12084.

Factor Structure and Etiological Architecture of the BRIEF: A Twin Study

Callie W. Little¹, Jeanette Taylor¹, Allison Moltisanti¹, Chelsea Ennis¹, Sara A. Hart^{1,2}, and Chris Schatschneider^{1,2}

¹Department of Psychology, Florida State University

²Florida Center for Reading Research, Florida State University

Abstract

Executive function is a broad construct that encompasses various processes involved in goal-directed behavior in non-routine situations (Banich, 2009). The present study uses a sample of 560 5- to 16-year-old twin pairs ($M = 11.14$, $SD = 2.53$): 219 MZ twin pairs (114 female; 105 male) and 341 DZ twin pairs (136 female, 107 male; 98 opposite-sex) to extend prior literature by providing information about the factor structure and the genetic and environmental architecture of the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000a, 2000b), a multi-faceted rating scale of everyday executive functions. Phenotypic results revealed a 9-scale 3-factor model best represents the BRIEF structure within the current sample. Results of the genetically sensitive analyses indicated the presence of rater bias/contrast effects for the Initiate, Working Memory and Task-Monitor scales. Additive genetic and non-shared environmental influences were present for the Initiate, Plan/Organize, Organization of Materials, Shift, Monitor and Self-Monitor scales. Influences on Emotional Control were solely environmental. Interestingly, the etiological architecture observed was similar to that of performance-based measures of executive function. This observed similarity provided additional evidence for the usefulness of the BRIEF as a measure of “everyday” executive function.

Executive function is a broad construct that encompasses various processes involved in goal-directed behavior in non-routine situations (Banich, 2009). Planning, initiating required behavior, inhibiting irrelevant behavior, maintaining attention, and other processes are aspects of executive functioning (Banich, 2009). Given its relevance to human behavior, executive function has been widely studied and found to be multifaceted with certain facets (e.g., attention) more strongly associated with certain problems (e.g., attention-deficit/hyperactivity disorders; (Biederman et al., 2004; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). As research over the past two decades has shown the association of executive function to various human behaviors and disorders, attempts have been made to reveal the etiological architecture of the facets of executive functioning using twin studies to disentangle the influence of genetic from environmental factors. This work has provided insights into the types of etiological factors that contribute to individual differences in

*Requests for reprints should be addressed to Callie W. Little, 1107 W. Call Street, Florida State University, Tallahassee, FL, 32306, USA; little@psy.fsu.edu.

certain facets of executive functioning. The present study extends prior literature by providing information about the genetic and environmental architecture of the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000a, 2000b), a multifaceted measure of executive functioning that is gaining popularity in clinical and research settings.

Executive functioning encompasses processes that begin to develop in childhood and continue through late adolescence (Anderson, 2002). Anderson (2002) noted that the assessment of executive functioning in children should be developmentally appropriate but also ecologically valid. Much of the literature on executive functioning has focused on performance-based tasks in which a facet (e.g., response inhibition) is assessed directly through a laboratory analog task (e.g., using a computer to present letters on a screen and asking a child to press or not press a keyboard key based on given conditions). Although there are clear advantages to such measures in that they can be designed to closely map onto a particular facet of executive functioning in a developmentally sensitive way, performance-based measures often lack ecological validity (Anderson, 2002). In contrast, the BRIEF was designed to measure “everyday” executive functioning by asking caregivers or teachers to rate children (age 5 to 18) on behavior that occurs in home and/or school settings (e.g., interrupts; fails to organize work; Gioia et al., 2000a). Its format provides an advantage for large-scale studies in which measures of executive functioning are needed but laboratory visits are prohibitive for assessment. In addition to the BRIEF, multiple rating-scales for measuring executive functioning have been developed such as the Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008) and the Children’s Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001). These scales also show supporting evidence as reliable and valid measures of executive functioning skills in young children, however, they are limited in comparison to the BRIEF in the age range that they have been determined to reliably measure and the depth and breadth of executive functioning skills in which they are able to assess. Therefore, the BRIEF is more suited to assess a wider range of executive functioning skills within the broad age range of the current sample.

The scale and factor structures of the BRIEF have been examined in several previous investigations using both clinical and non-clinical samples; however, there are inconsistencies among the findings. As originally conceptualized, the BRIEF is comprised of eight facet scales that combine to form two theoretically-derived index scales and a global index scale (Gioia et al., 2000a, 2000b). The Behavioral Regulation Index (BRI) is the sum of the Inhibit (stopping a response), Shift (enacting and accepting change); and Emotional Control (controlling emotions) scale scores. The Metacognition Index (MI) is the sum of the remaining facet scores: Initiate (getting started on a task without supervision), Working Memory (remembering things), Plan/Organize (planning and organizing behavior and thoughts), Organization of Materials (messiness in various settings and situations), and Monitor (monitoring behavior). The Global Executive Composite (GEC) is the sum of all eight facet scale scores. The BRIEF manual presents data showing concurrent and discriminant validity of the scales (Gioia et al., 2000b). Other published studies have shown that the BRIEF differentiates children with ADHD (Gioia, Isquith, Kenworthy, & Barton, 2002; Mahone et al., 2002; McAuley, Chen, Goos, Schachar, & Crosbie, 2010), reading disability (Gioia, Isquith, Kenworthy, et al., 2002), autistic spectrum disorder (Gioia, Isquith,

Kenworthy, et al., 2002), and severe traumatic brain injury (Gioia, Isquith, Kenworthy, et al., 2002) from controls. There is also evidence that the BRIEF is a better predictor of ADHD status than performance-based measures like the stop-signal task (Toplak, Bucciarelli, Jain, & Tannock, 2008). Thus, the BRIEF appears to provide an assessment of multiple facets of executive functioning in children and has predictive value in discriminating those with and without disorders that are marked by executive functioning deficits such as ADHD. However, questions remain about the factor structure of the BRIEF that should be addressed to enhance confidence in and utility of the measure.

The factor structure of the BRIEF has been examined primarily in clinical samples. In the first of these studies (Gioia, Isquith, Retzlaff, & Espy, 2002), the authors of the measure used parent-rated data from a sample of 374 children age 5–18 with ADHD, learning disorder, autism spectrum disorder, Tourette's, or seizure disorder to conduct a confirmatory factor analysis (CFA) not of the original 8-scale, 2-factor model they outlined in the manual but rather a model in which the Monitor scale was separated into Task-Monitor and Self-Monitor based on their earlier re-analysis of that scale. They found that a 3-factor model provided the best fit: Metacognition (Initiate, Working Memory, Plan/Organize, Organization of Materials, and Task-Monitor); Emotional Regulation (Shift; Emotional Control); and Behavioral Regulation (Self-Monitor; Inhibit). Egeland and Fallmyr (2010) compared the original 8-scale model to the 9-scale model in a Norwegian sample of 158 9- to 11-year-olds drawn from clinical referrals or schools and concluded that the 3-factor model fit best for parent- and teacher-rated data, however, the fit indices suggested that all models tested provided similar fit. In an even smaller study of 80 children with epilepsy, Slick, Lautzenhiser, Sherman, and Eyrl (2006) conducted an exploratory factor analysis using parent rated data and concluded that the original 8-scale, 2-factor model was supported, but the results appeared to support a unitary factor model as well. Donders, DenBraber, and Vos (2010) also found support for the original 2-factor model using exploratory factor analysis on data from a sample of 100 parents rating their child with a traumatic brain injury. Studies using clinical samples have failed to converge on a clear factor solution for the BRIEF, and the authors of the measure called for studies in non-clinical samples (Gioia, Isquith, Retzlaff, et al., 2002) which could provide greater power but also identify differences between clinical and normally developing children.

The only published study found on the factor structure of the BRIEF in a non-clinical sample was conducted with parents of 847 Dutch children age 5–18 years (Huizinga & Smidts, 2010). The BRIEF was administered twice within a 6-week interval and the data were divided into four age bands to assess age-related differences in executive functioning. Huizinga and Smidts examined only the original 8-scale, 2-factor model and it fit the data across all four age bands. Notably, this study also provided confirmatory evidence of the internal consistency of the BRIEF scales (all above .77) as well as evidence of the reliability of the parent-rated scale scores with most test-retest correlations in the .8 to .9 range (Working Memory was lowest but still good at .73). The Huizinga and Smidts study provided confidence in the original structure outlined for the BRIEF but failed to offer a comparison to the revised 9-scale, 3-factor structure identified by the authors of the measure (Gioia, Isquith, Retzlaff, et al., 2002). Further clarification of the underlying factor structure of the BRIEF can provide evidence for whether the theoretical constructs represented can be

meaningfully separated into one or multiple areas of executive functioning. Evidence supporting a three-factor model as the best fitting model will support existing theories of executive functioning as a multifaceted construct (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Rabbitt, Lowe, & Shilling, 2001). If executive functioning can be conceptualized as multifaceted, as suggested by the three factor model, this has important implications for understanding the cognitive operations and behaviors associated with underlying executive functioning skills as well as treatment planning for disorders influenced by deficits within specific subfacets of executive functioning. The present study will fill the gap in the literature by using CFA to compare the original (8-scale) and the revised (9-scale) models in a large non-clinical sample in order to help clarify the factor structure of the BRIEF.

While the factor structure of the BRIEF lacks clarity based on the extant literature, the etiological architecture of the measure is completely unknown. Twin studies provide a means of estimating genetic and environmental influence on a behavior or trait by comparing the similarity of monozygotic (MZ) or identical and dizygotic (DZ) or fraternal twins which differ naturally in their proportion of shared genes with MZ twins sharing all copies of their genes and DZ twins sharing, on average, 50% of segregating genes. Such samples are used to estimate the amount of the variation in a behavior or trait that is associated with latent factors representing additive genetic effects (A) that serve to make biologically related people similar, shared environmental effects (C) that serve to increase similarity among people who share an environment; and non-shared environmental effects (E) that serve to differentiate people from one another. Non-additive genetic effects such as dominance (D) can also be estimated and they represent influences of particular combinations of genes. Although genetically informative studies using twin pairs have not yet been conducted on the BRIEF, they have been used to examine performance-based measures of executive functioning.

Twin studies of performance-based measures of executive functioning suggest that additive genetic influence is substantial for many (Anokhin, Heath, & Ralano, 2003; Friedman et al., 2008; Lee et al., 2012; Polderman et al., 2006; Taylor, 2007; Vasilopoulos et al., 2012) although not all (i.e., Wisconsin Card Sorting Test; Kremen, Eisen, Tsuang, & Lyons, 2007; Taylor, 2007). Two studies found support for a common pathway etiological model in which a common additive genetic factor accounted for a significant amount of the genetic influence on multiple executive functioning tests (Friedman et al., 2008; Lee et al., 2012), indicating that individual differences in various facets of executive functioning stem from the same genetic factors. None of the prior studies found evidence of shared environmental influences on performance-based measures of executive functioning and, instead, all found evidence of non-shared environment. Thus, the literature on the etiological influences underlying performance-based measures of executive functioning shows that most are influenced by additive genetic and non-shared environmental factors. No prior study has addressed the question of etiological influences on the BRIEF and the present study will fill this gap in the literature using parent-rated data on a large sample of twins.

The BRIEF provides an assessment of everyday executive behaviors (Gioia et al., 2000a) and it is used in research and clinical settings. Few studies have examined the factor structure of the BRIEF and only one study (Huizinga & Smidts, 2010) has done so in a non-

clinical sample, but that study examined only the original 8-scale, 2-factor structure and results from clinical samples suggest that a 9-scale, 3-factor structure may be more appropriate (Gioia, Isquith, Retzlaff, et al., 2002). The first aim of this study was to compare the original 8-scale structure of the BRIEF to the revised 9-scale structure in 1-, 2-, and 3-factor models in a large non-clinical sample of children. This will provide an empirical test of all of the previously investigated structures from clinical and non-clinical samples. It is hypothesized that both the 2- and 3-factor structures will fit the data better than a 1-factor model based on prior factor analytic work on the BRIEF. However, the mixed results in the prior literature did not allow predictions about the relative performance of the multi-factor models to be made with confidence. The second aim of this study was to take advantage of the genetically informative nature of the sample and conduct an initial examination of the etiological architecture of the BRIEF.

Method

Participants

Twins for this study were recruited through the Florida State Twin Registry (Taylor, Hart, Mikolajewski, & Schatschneider, 2012; Taylor, James, Reeves, Bobadilla, 2006) as participants for the Florida Twin Project on Reading (Taylor & Schatschneider, 2010). The project utilizes achievement data from the Progress Monitoring and Reporting Network, a statewide educational database, as well as data ascertained through packets sent to the families during the 2012 calendar year. These packets included a letter about the study, consent and assent forms, a parent questionnaire, child questionnaires (for twins age 9 and older), and a five-question assessment of zygosity (Lykken, Bouchard, McGue, & Tellegen, 1990) for families who were new to the Registry.

Families that were already part of the Registry received a study packet directly through the mail. Potential new families with twins in kindergarten through third grade were recruited by sending packets home through schools following the approach used previously (Taylor & Schatschneider, 2010). Of the 3,343 packets sent out, 568 families participated, 106 refused, 2334 did not respond and 335 were not reached either because they had moved and could not be located, the school principal refused to participate (i.e., packets were not sent home with potential new twins), or the children were not twins. Three families did not report zygosity information and were not included in the analyses. An additional 5 families were excluded because the parent rating of the BRIEF was invalid based on scoring criteria in the manual. The final sample included 560 5- to 16-year-old twin pairs ($M = 11.14$, $SD = 2.53$). Table 1 presents the age distribution of the sample. Of the twin pairs, 219 were monozygotic (MZ; 114 female; 105 male) and 341 were dizygotic (DZ; 136 female, 107 male; 98 opposite-sex). Parent-reported racial composition of the sample was 72% White, 13% African American, 8% mixed race, 2% Asian, 4% Other and 1% did not provide information on race. Ethnic composition for this sample was 28% Hispanic and 69% non-Hispanic, with 3% failing to report ethnicity. Parents reported household income on a scale ranging from 1 "less than \$10,000" to 12 "\$210,000 or more." The mean self-reported household income for the sample was 4.88 ($SD = 2.83$), which corresponds roughly to the "\$70,000 to \$89,000" option).

Procedure and Measure

Instructions sent with a packet of questionnaires asked that they be completed by a rearing parent or legal guardian of the twins who knew them well (in most cases the biological mother completed the questionnaires). The parent packet was organized such that all measures were rated on a designated member of the twin pair and then a second copy of the measures was rated on the co-twin. This was intended to help decrease similarity or contrast effects that can occur when twins are rated simultaneously and by a single rater as with the questionnaires in the present sample. Parents returned the completed consent form, assent forms, and questionnaire packet in a pre-paid envelope and then they were sent a gift card to a retailer of their choice.

Behavior Rating Inventory of Executive Function – Parent Form—The BRIEF (Gioia et al., 2000b) is comprised of 86 items assessing behaviors exhibited by the child in the past 6 months. All items are rated on a 3-point scale (1= Never, 2= Sometimes, 3= Often), and items are worded in the pathological direction such that higher scores indicate more problems in an area. The BRIEF manual outlines eight clinical scales, two index scales, and one composite index scale (Gioia et al., 2000). The clinical scales include the Inhibit scale (10 items) that measures problems resisting impulses. The Shift scale (8 items) measures problems in flexibly changing behavior and making transitions. The Emotional Control scale (10 items) assesses problems modulating emotional reactions. The Initiate scale (8 items) assesses problems starting tasks and identifying strategies or responses. The Working Memory scale (10 items) measures problems storing information with the purpose of completing a task. The Plan/Organize scale (12 items) measures problems planning for future events, setting goals, and developing strategies. The Organization of Materials scale (6 items) measures problems related to orderliness of work, play, and storage spaces. The final clinical scale, Monitor (8 items) measures problems related to work-checking habits such as failing to check work after completing it.

Sum scores were computed for each scale. Previous research with the BRIEF has demonstrated strong psychometric properties including high internal consistency ($\alpha = .80$ to $.98$) and good test-retest reliability over a 2-week period ($r = .81$). In the current investigation, the clinical scales showed good internal consistency (alpha reliability ranged from $.83$ to $.93$). Following the procedure in prior factor analytic studies (Gioia, Isquith, Retzlaff, et al., 2002; Huizinga & Smidts, 2010), raw rather than standard scale scores were used in analyses.

Analyses

Huizinga and Smidts (2010) found age effects on the BRIEF scales, but the present sample was not large enough to create similar age bands. Moreover, the present sample of twins was also used to examine genetic and environmental influences on the BRIEF and estimates of these influences can be biased when twin pairs of varying ages are used. As such, the raw scale data were corrected for age as outlined by McGue and Bouchard (1984) by regressing age and age-squared onto each variable, and those age-corrected scores were used in all analyses.

To address the first aim of the study, six CFA models were fit to BRIEF scale data using the ML estimator in Mplus 7.11 (Muthén & Muthén, 1998–2012). Three models utilized the original eight scales from the BRIEF: a 1-factor model with all eight scales as indicators; the original 2-factor model as described in the introduction above; and a novel 3-factor model designed to mirror the 3-factor solution obtained by Gioia, Isquith, Retzlaff, et al., (2002). Specifically, the 8-scale, 3-factor model included: Metacognition (Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor as indicators), Emotional Regulation (Shift; Emotional Control as indicators), and Behavioral Regulation (with Inhibit as the sole indicator modeled as a latent variable by using $[1 - \text{scale reliability}] * \text{scale variance}$ in specifying a known amount of measurement error as outlined by Bollen, 1989. Using a single indicator prevents identification of the model because it does not allow the unique variance for this indicator to be estimated. Bollen's method fixes the unique variance, allowing the latent variable to be specified and the model identified when only one indicator is available. The other three CFA models utilized the 9-scale scoring of the BRIEF in which the Monitor scale is scored as two scales (Task-Monitor and Self-Monitor). The 1-factor model included all nine scales as indicators. A novel 9-scale, 2-factor model was designed to test the original two factors for the BRIEF and included Behavioral Regulation (Inhibit, Shift, Emotional Control, Self-Monitor) and Metacognition (Initiate, Working Memory, Plan/Organize, Organization of Materials, and Task-Monitor). Finally, the 9-scale, 3-factor solution obtained by Gioia, Isquith, Retzlaff, et al., (2002) and described in the introduction above was modeled. The fit of each model was evaluated using multiple criteria: the chi-square statistic, Akaike's Information Criterion (AIC; Akaike, 1987), the root mean square of approximation (RMSEA) and Bentler's Comparative Fit Index (CFI; Hu & Bentler, 1999). Chi-square values closest to the degrees of freedom indicate a better fitting model. Akaike's Information Criterion is a modified version of chi-square (taking into account model complexity) thus with both chi-square and AIC indices, lower values indicate a better fitting model. Values of the CFI above .95 indicate close model fit, whereas, for the RMSEA, values less than .08 indicate adequate model fit (Browne & Cudeck, 1993).

To address the second aim of the study, intraclass correlations were calculated by zygoty for all BRIEF scales using the age-corrected scores. Given that MZ twins share 100% of their segregating genes and DZ twins share, on average, 50% of their segregating genes, twin intraclass correlations indicate the within-pair similarity and offer a simple means of inferring genetic and environmental influence. When the MZ correlation is twice the magnitude of the DZ correlation, then additive genetic (A) effects are inferred. If the MZ correlation is more than twice the magnitude of the DZ correlation, then dominance genetic (D) effects are inferred. When the MZ correlation is less than twice the magnitude of the DZ correlation, then shared environmental (C) effects are inferred. Finally, non-shared environment (E) is inferred when the MZ correlation is less than 1. Note that the E estimate also includes measurement error.

The inclusion of opposite-sex DZ twins allowed for an evaluation of sex differences in the etiological factors associated with BRIEF scales. When the opposite-sex DZ correlation is less than the average of the same-sex DZ correlations, then qualitative sex differences may exist such that different etiological factors affect individual differences in boys and girls on

the trait. Quantitative sex differences may exist if the same-sex DZ correlations differ in magnitude between males and females.

In addition to sex differences, rater bias and sibling interaction may influence estimates of genetic influence through their effect on the DZ correlation. Genetic influence is estimated when the DZ correlation is lowered relative to the MZ correlation. This can occur through negative sibling interaction in which siblings differentiate from one another by behaving in opposing ways on a trait. Rater bias in the form of contrasting the behavior of twins based on their perceived zygosity also serves to lower the DZ correlation. Rater bias and negative sibling interaction may produce what appear to be D effects. Positive sibling interaction wherein siblings cooperate on a behavior serves to increase the DZ correlation, which will result in the appearance of C effects. Previous behavioral genetic studies have found evidence for the presence of rater bias and/or contrast effects when using survey-based measures of behavior (i.e. ADHD) that rely on parent report (Eaves et al., 1997; Ebejer et al., 2014; Merwood et al., 2013; Nadder, Silberg, Eaves, Maes, & Meyer, 1998; Simonoff et al., 1998), therefore, when examining the etiology of survey-based measures of behavior it is important to test for potential presence of rater bias or contrast effects that may lead to over or under estimating genetic influences. Contrast effects act in the same manner as dominance genetic influences, therefore they cannot be estimated in the same model where C is estimated. Evidence for negative sibling interaction or rater effects in ADHD and other behaviors related to executive functioning have been found within previous literature, suggesting that the MZ correlations greater than two times the DZ correlations found in the current sample may represent either dominance genetic influences or potential rater bias/sibling interaction. Since no previous evidence exists for the presence of sibling cooperation effects for these related behaviors the b parameter was not estimated in models including C.

The twin correlations provide insights into which biometric models to fit to the data in order to obtain robust estimates of genetic and environmental influences. Phenotypic variance (V_P) in a trait or behavior is comprised of the variance (V) among the aforementioned etiological components: additive genetic (A), shared environmental (C), dominance genetic (D) and non-shared environmental (E) influences.

$$(1) \quad V_P = V_A + V_C + V_D + V_E$$

However, twin data are not sufficient to model C and D effects simultaneously and, therefore, separate models are needed ($A + C + E$ or $A + D + E$). Biometric models are specified by supplying the expected covariance among twins as follows (where D is fixed to zero when fitting the ACE model and C is fixed to zero when fitting the ADE model):

$$(2) \quad \text{MZ covariance} = A + D + C$$

$$(3) \quad \text{DZ covariance} = .5A + .25D + C.$$

With data from only one rater (as in the present study), it is not possible to differentiate rater bias from negative sibling interaction. However, it is possible to estimate a rater bias/sibling interaction parameter, b , in structural equation models by specifying a reciprocal path between the observed phenotypes within pairs (see Figure 2). The b parameter is estimated from the differences between the total variance of MZ and DZ twins and with sibling

interaction/rater bias effects the b parameter represents the influence of one twin's phenotype on the other or one parent's rating of one twin on their rating of the other (Simonoff et al., 1998). Sibling interaction or rater bias is inferred from significant negative values of b , where the influence of one twin or rating leads to behaviors or ratings in the opposite direction for the co-twin. Positive interaction or sibling cooperation is inferred from positive values, where the influence of one twin leads to more behaviors in the same direction as their co-twin. Models were fit to age-corrected scale data using full information maximum-likelihood estimation in Mx (Neale, Boker, Xie, & Maes, 2006). These models produce estimates of the standardized variance for each parameter with 95% confidence intervals (CIs) that indicate significance when they do not include zero.

In addition to the univariate models for each individual BRIEF scale, a model can be fit that estimates A, C, and E (or A, D, and E) on the latent factors from the CFA models. As outlined below, the results from the twin correlations and univariate models suggested the presence of C on some scales and the presence of D (or rater bias/sibling interaction) on others – some of which were part of the same latent factor in the CFA models (i.e. Behavioral Regulation, Emotional Regulation and Metacognition). As noted above, twin data are not sufficient to estimate C and D effects simultaneously and, as such, the present study provides data on the etiology of only the individual scales of the BRIEF (i.e. Shift, Inhibit, Emotional Control, etc.).

Results

Prior to any analyses, distributions of the scales were examined and checked for outliers. Outliers for each scale were removed by assigning missing values to scores above or below three standard deviations from the mean. None of the distributions required transformation.

Factor Structure of the BRIEF

The upper part of Table 2 presents a summary of the six CFA models that were fit to age-corrected scale scores. If the three executive functions suggested by the 3-factor model are separate constructs from each other, then the three factor models should provide the best model fit to the data. Furthermore, if the original Monitor subscale is better represented as two distinguishable constructs of task- and self-monitoring, then the 9-scale models should provide a better fit to the data than the 8-scale models. As hypothesized, the 2- and 3-factor structures fit the data better than a 1-factor model as indicated by significant chi-square difference tests, lower AIC and RMSEA values and higher CFI values for the multi-factor models. The 9-scale, 3-factor model was deemed the best fitting model based on multiple goodness-of-fit indicators. As shown in Table 1, between the 8 and 9 scale models, the chi-square statistic was low and close to the degrees of freedom for these models. Although the chi-square difference test was not significant between the 8-scale, 3-factor and the 9-scale, 3-factor models, the 9-scale, 3-factor model was chosen as the best representation of the data based on its parsimony and the values of additional fit statistics. Within the 9-scale models, both the two and three factor models showed significantly better fit to the data based on chi-square difference tests. The three factor model showed the best fit to the data based on the

lowest AIC value, a CFI estimate of above .95 and RMSEA closest to, but not lower than .08. The 9-scale, 3 factor model is presented in Figure 1.

Means and SDs for each BRIEF scale, as suggested by the CFA model results, are provided in Table 3. To allow comparisons with findings in the literature, raw scale scores are presented. In addition, the original Monitor scale as well as the separate Task-Monitor and Self-Monitor scales are included to provide maximal information about the BRIEF in the present sample. Means and standard deviations for males and females were largely comparable to those of the normative sample for the BRIEF (10.38–19.53; Gioia et al., 2000b) and the Dutch community sample used by Huizinga and Smidts (2010; 12.6–21.7), although some scale means were somewhat lower for the present sample. This suggests that twins are fairly representative of the general population (an assumption in twin research) with regard to BRIEF scales.

Huizinga and Smidts (2010) found gender differences in the factor scores of the BRIEF. As such, gender differences on the scales were examined in the present sample using independent samples *t*-test and Levene's Test for Equality of Variances. To avoid inflated sample size and to meet the assumption of independent observations, one twin was randomly selected from each pair for the analysis¹. Table 4 presents the comparisons using age-corrected scale scores. The majority of scales showed a mean and/or variance difference with males having more problems in a given area and/or greater variability. Based on these findings, the presence of measurement invariance across genders was examined for using the original best-fitting CFA model. The 9-scale, 3-factor model was examined for gender differences in its structure by first fitting a baseline model with factor loadings free to vary across gender. That model was compared to one that constrained the factor loadings to be equal across gender (Model 1). Next, stricter forms of invariance were assessed by constraining factor variances and covariances across genders (Model 2). A comparison of model fit indices revealed no significant difference in model fit between the baseline model and models 1 ($\chi^2(6)=11.00$ $p>.05$) or 2 ($\chi^2(6)=4.46$ $p>.05$), suggesting the BRIEF constructs were not functioning differently across genders.

Genetic and Environmental Influence on the BRIEF

An important assumption in twin research is that MZ and DZ twins have equally similar environments which allows for increased MZ similarity, when present, to be attributed to genetic factors rather than to more similar treatment in the environment. In order to check this assumption, means and variances of MZ and DZ twins were compared for each scale within each sex using the same procedure outlined above in testing for gender differences (on the same subsample). As can be seen in Table 5, there were no mean differences and only three variance differences between MZ and DZ twins (all among females), indicating that the equal environments assumption likely holds for BRIEF scales. However, these variance differences could also signal rater bias/sibling interaction effects which can mask as C or D effects, therefore, their presence is explored within the present analyses, as well.

¹Phenotypic analyses were also conducted with the other member of each twin pair randomly selected. These results revealed nearly identical patterns of model fit statistics and parameter estimates, suggesting no systematic differences across members of the twin pairs.

Given the mean gender differences found for the BRIEF scales, twin correlations were calculated by gender and are presented in Table 6. The significance of each correlation is shown along with the z test comparing the MZ and DZ correlations within gender. Several scales (Shift, Emotional Control, Monitor, and Self-Monitor) showed a pattern of correlations suggestive of C influences for both genders, indicated by MZ correlations less than two times DZ correlations for these scales. Emotional Control showed no difference in correlation magnitudes across zygosity for either gender indicating no genetic influence for that scale. All other scales (except Shift for boys and Emotional Control) evidenced significantly greater MZ than DZ correlations, indicating genetic influence. Task-Monitor showed a pattern of correlations suggestive of D influences for both genders, indicated by MZ correlations greater than two times correlations for DZ twins. The remaining scales showed inconsistent patterns across gender. The opposite-sex DZ correlation was lower than the average same-sex DZ correlation for all scales, suggesting the possibility of sex-specific effects of genes or environment. Moreover, two of the female DZ correlations were non-significant and near zero, suggesting negative sibling interaction or rater bias. Biometric models included tests of sex-limitation for scales in which the same model was fit across gender. Rater bias/sibling contrast effects were estimated for all scales in which D influences were indicated.

The biometric model-fitting results are summarized in Table 7. For these analyses, opposite-sex twins were ordered with males as twin 1 and females as twin 2. For scales in which dominance genetic influences or rater bias/sibling interaction were indicated by the intraclass correlations a model estimating additive genetic, dominance genetic, non-shared environmental and rater bias/sibling interaction influences (ADE-B) was first fitted to the data. Next, submodels were compared to the ADE-B model using chi-square difference tests. First, a sub-model in which the b estimate was dropped (ADE) was compared to the original ADE-B model, then a model in which the d estimate was dropped (AE-B) was compared to the original model, finally, a submodel in which both d and b estimates were dropped (AE) was compared to the AE-B model. For scales in which shared environmental influences were indicated by the intraclass correlations an initial model including estimates of additive genetic, shared environmental and non-shared environmental influences (ACE) was fitted to the data, then two submodels (AE and CE) were fitted and compared using chi-square difference tests to the parent ACE model. Model fit estimates, and parameter estimates for additive genetic (a), dominance genetic (d), shared environmental (c), non-shared environmental (e) and rater bias/sibling interacts (b) with confidence intervals were listed in columns 2–9 for each model and submodel tested within all scales of the BRIEF. Chi-square difference tests with degrees of freedom comparing constrained and unconstrained models across gender and within scale are presented in column 10. When the unconstrained model was the better fitting model, indicating potential gender differences, parameter estimates were indicated for both genders. Column 11 presents the chi-square difference tests between models and submodels for each scale. Results indicated significant and moderate (.34–.73) genetic effects for all scales except Emotional Control and smaller (most around .30) but significant environmental effects for all scales. Significant, negative b parameters were estimated for females, but not males on the Inhibit, Task-Monitor, Organize and Working Memory scales (–.18–-.22), indicating negative sibling interaction or rater

bias. Additionally, D estimates for these scales were low and non-significant for both genders (.00–.18) and dropping D from the models did not result in significantly poorer model fit. An AE-B model was indicated as the best fitting model for Inhibit, Working Memory, and Task-Monitor for females and males. Tests of heterogeneity for these subscales revealed significant χ^2 differences when all parameters were constrained across genders. When Monitor was scored as a single scale, the twin correlations suggested an AE model would fit for both genders. Similarly, the Plan/Organize, Self-Monitor, Initiate, Shift, and Organization of Materials scales were associated with A and E effects for both boys and girls. Tests of heterogeneity for the Organize scale indicated no significant reduction in model fit when estimates were constrained across gender suggesting no gender differences were present for this scale. Finally, Emotional Control was the only scale to show significant C and non-significant A effects and this pattern of influences was present for both genders.

Discussion

The BRIEF is a measure of everyday executive functioning that is multifactorial in design and captures a wide spectrum of behavior from planning and organization to working memory to emotional control. As such, the BRIEF has the potential to provide a means of assessing a full range of executive functioning in clinical and non-clinical samples alike. The present study used a genetically informative community sample of twins to address the inconsistency in the literature on the factor structure of the BRIEF and to provide initial data on the extent to which variance in individual BRIEF scale scores is associated with genetic and environmental factors.

Prior studies of the factor structure of the BRIEF come primarily from clinical samples using exploratory factor analysis. The only published study using a non-clinical sample and CFA methodology examined just the original 8-scale, 2-factor solution and found that it fit the data for multiple age groups within their sample (Huizinga & Smids, 2010). The present findings extend the literature by showing that the 9-scale, 3-factor structure derived empirically in a clinical sample (Gioia, Isquith, Retzlaff, et al., 2002) provided a better fit than the original 8-scale, 2-factor structure, unitary structure, and even a novel structure (8-scale, 3-factor) in a community sample. The examination of 1-, 2-, and 3-factor structures for both the 8- and 9-scale scoring of the BRIEF demonstrated that the BRIEF is clearly multi-factorial and, indeed, the 3-factor structures fit the data best whether the 8- or 9-scale scoring was employed. Thus, it appears important to distinguish a Behavioral Regulation from an Emotional Regulation factor whether using the original 8 scales of the BRIEF or the 9-scale scoring in which Monitor is separated into Self- and Task- components.

It is not certain that findings from non-clinical samples should correspond closely to those from clinical samples given that the severity and nature of executive functioning problems is greater in the latter and could produce a different structure. With this caveat in mind, it is notable that the results of this study are consistent with those of the only other investigation that compared the original 8-scale, 2-factor model to the 9-scale, 3-factor model in a mixed clinic-referred and non-clinical sample (Egeland & Fallmyr, 2010). That study used parent- and teacher-rated data and the 9-scale, 3-factor model was favored for each. When taken together with the CFA results of Gioia, Isquith, Retzlaff, et al. (2002) using parent-rated data

on a clinical sample, it appears that the structure of the BRIEF – whether rated by teachers or parents on children from the community or referred by clinics – is best characterized by nine scales forming three factors: Behavioral Regulation, Emotional Regulation, and Metacognition. Moreover, the present study provides evidence that this structure is invariant across gender despite the fact that boys show significantly higher means on some BRIEF scales as compared to girls in this and other samples (Gioia et al., 2000b; Huizinga & Smidts, 2010).

The presence of gender differences was not evident in the structure of the BRIEF, and results of further investigation of gender-specific etiological architectures for BRIEF scales were consistent with these findings. For all scales, the same models could be fitted for males and females with some differences between significance for parameter estimates (i.e. estimates significant for females, but not males). Inhibit and Working Memory are the two BRIEF scales that are most closely related to features of ADHD (Gioia et al., 2000b), which often indicates strong genetic influence in the literature (Nikolas & Burt, 2010; Rietveld, Hudziak, Bartels, Van Beijsterveldt, & Boomsma, 2003). Inhibit and Working Memory scales indicated the presence of additive genetic influences as well as specific environmental influences and contrast effects or rater bias, consistent with findings from ADHD literature (Ebejer et al., 2014; Nadder, Silberg, Eaves, Maes, & Meyer, 1998; Simonoff et al., 1998). Additionally, the Task-Monitor scale, which includes items such as checking homework that could be influenced by ADHD behaviors (Power, Werba, Watkins, Angelucci, & Eiraldi, 2006) also showed the presence of contrast effects or rater bias. Although the AE-B model was the best fitting model for both genders for these scales, the contrast effect/rater bias parameter was significant only for female twins. Because only one rater was used in the current study, it was not possible to determine whether these influences were due to rater bias or contrast effects and, therefore, beyond the current scope to suggest potential causes of these gender differences. Parent surveys were organized such that the rating parent filled out all items of the BRIEF for one twin, then again for the co-twin. This type of organization may have negated rater bias by reducing the potential influence of one twin's rating on his or her co-twin more so than asking the rater to fill out each item for both twins simultaneously.

The predominant influences for the Shift, Initiate, Plan/Organize, Monitor, Self-Monitor and Organize scales were additive genetic and specific environmental (AE model). These results are consistent with previous literature on performance-based measures of executive functioning which indicated negligible shared environmental influences (Friedman et al., 2008; Malone & Iacono, 2002) and provide evidence that the BRIEF scales show similar etiological architecture as performance-based measures.

Influences for Emotional Control were limited to shared and non-shared environmental effects. A previous genetically sensitive examination has found evidence that shared and non-shared environment are the primary influences on negative emotionality (Neiss, Stevenson, Legrand, Iacono, & Sedikides, 2009) and, phenotypically, there is evidence that deficits in the home environment can negatively impact the development of emotionality and emotional control, consistent with the current results (Rekart, Mineka, Zinbarg, & Griffith, 2007).

The differences in the relative influence of genetic and environmental influences across BRIEF scales present some challenges for testing genetically sensitive models of the BRIEF's phenotypic structure. The 9-scale, 2-factor structure favored in the present CFA results could not be examined using twin data alone as A, C, D, and E cannot be modeled simultaneously with only twin data, and all of these factors were important to at least one BRIEF scale. Users of the BRIEF can utilize higher order factors such as the Behavioral Regulation Index or the Metacognition Index, but users should bear in mind that the scales that comprise the given index may not stem from the same etiological influences.

It is important to consider the limitations of the current study when interpreting its findings. The large age range of the sample may have resulted in age-related differences on performance across the BRIEF subscales. In previous studies, executive functioning skills have been found to increase throughout childhood and adolescence (Huizinga, Dolan, & van der Molen, 2006; Huizinga & Smidts, 2010). Additionally, an investigation of the factor structure of executive functioning has produced results indicating that the factor structure remains invariant across ages 5–18 (Huizinga et al., 2006), suggesting that the increase in executive functioning skills during childhood and adolescence may not directly impact the subcomponent arrangement. However, the current sample was not large enough to divide into age bands and test for potential age-related differences, specifically. Therefore, the data were age-corrected for the present analyses. Future investigations of the factor structure and the etiology of the BRIEF may benefit from looking across age bands where possible. Additionally, values less than .08 are preferred for RMSEA statistics, however, the estimated RMSEA value within the best fitting model was .09 (CI: .07–.10), suggesting that although this model best represented the current data the overall model fit was not ideal. Moreover, the confidence interval around this RMSEA value fell below .08, indicating that the estimated value should be interpreted cautiously. With only one rater for the measure, the *b* effects could not be established as either rater bias or contrast effects; therefore, it is unclear which of these was driving the low DZ female correlations for some of the scales. Items on the BRIEF are mixed (i.e., not grouped by scale), and the correlations for DZ twins were not low for all scales, suggesting measurement error or rater bias is less likely. A future direction will be to investigate the reliability of variance across different time points of the BRIEF administration once additional rounds of survey data have been collected. The present investigation included ratings from the parents of the twins. Alternatively, teachers could have been chosen to rate the twins on this measure since they have knowledge and insight into how their students behave in classroom settings for several hours during the day. However, having the parent rate the behavior of the twins allowed us to also obtain ratings on the home environment, which was necessary for addressing major aims of the broader study on reading. Future investigations can benefit from including two sets of raters (both parents and teachers) which will allow for clearer establishment of whether rater bias or contrast effects are present. The present examination provided evidence that a 9-scale, 3-factor model best represents the BRIEF structure within a non-clinical sample, suggesting that this may be a more optimal structure to use than the one provided in the manual for the measure. This study provides the first evidence of the etiological factors that underlie individual differences in BRIEF scores. Understanding the etiological influences on subscales of the BRIEF has important implications for knowing not just what skills are being

measured, but how these skills are being impacted. Differences and similarities in the etiological influences across subscales can be taken into consideration when developing interventions targeted to skills measured by the BRIEF, particularly for groups performing at atypical levels. Furthermore, the fact that the BRIEF shows similar types of etiological influences on its scales as performance-based measures of executive function provides additional confidence in this measure of “everyday” executive function.

Acknowledgments

The first author was supported by Predoctoral Interdisciplinary Fellowship (funded by the Institute of Education Sciences, US Department of Education (186000-520-025833). The research project was supported, in part, by a grant from the National Institute of Child Health and Human Development (P50 HD052120). Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies. The authors wish to thank the twins and their families for their participation in making this research possible.

References

- Akaike H. Factor analysis and aic. *Psychometrika*. 1987; 52:317–332.
- Anderson P. Assessment and development of executive function (EF) during childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2002; 8:71–82. [PubMed: 12638061]
- Anokhin AP, Heath AC, Ralano A. Genetic influences on frontal brain function: WCST performance in twins. *Neuroreport*. 2003; 14:1975–1978. [PubMed: 14561932]
- Banich MT. Executive function: The search for an integrated account. *Current Directions in Psychological Science*. 2009; 18:89–94.
- Biederman J, Monuteaux MC, Doyle AE, Seidman LJ, Wilens TE, Ferrero F, Faraone SV. Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *Journal of Consulting and Clinical Psychology*. 2004; 72(5):757. [PubMed: 15482034]
- Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.
- Browne MW, Cudeck R, Bollen KA. Alternative ways of assessing model fit. *Sage Focus Editions*. 1993; 154:136–136.
- Donders J, DenBraber D, Vos L. Construct and criterion validity of the Behaviour Rating Inventory of Executive Function (BRIEF) in children referred for neuropsychological assessment after paediatric traumatic brain injury. *Journal of Neuropsychology*. 2010; 4:197–209. [PubMed: 19930791]
- Eaves LJ, Silberg JL, Meyer JM, Maes HH, Simonoff E, Pickles A, Truett KR. Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry*. 1997; 38(8):965–980. [PubMed: 9413795]
- Ebejer J, Medland S, van der Werf J, Wright M, Henders A, Gillespie N, Duffy D. Contrast Effects and Sex Influence Maternal and Self-Report Dimensional Measures of Attention-Deficit Hyperactivity Disorder. *Behavior Genetics*. 2014:1–16.
- Egeland J, Fallmyr Ø. Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF): Support for a distinction between emotional and behavioral regulation. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2010; 16:326–337. [PubMed: 20209415]
- Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP, Hewitt JK. Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*. 2008; 137:201–225. [PubMed: 18473654]
- Gioia GA, Isquith PK, Guy SC, Kenworthy L. TEST REVIEW Behavior Rating Inventory of Executive Function. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2000a; 6:235–238. [PubMed: 11419452]

- Gioia, GA., Isquith, PK., Guy, SC., Kenworthy, L. Behavior Rating Inventory of Executive Function. Lutz, FL: Psychological Assessment Resources, Inc; 2000b.
- Gioia GA, Isquith PK, Kenworthy L, Barton RM. Profiles of everyday executive function in acquired and developmental disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2002; 6:121–137. [PubMed: 12638065]
- 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: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2002; 8:249–257. [PubMed: 12759822]
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
- Huizinga M, Dolan CV, van der Molen MW. Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*. 2006; 44(11):2017–2036. [PubMed: 16527316]
- Huizinga M, Smidts DP. Age-related changes in executive function: A normative study with the Dutch version of the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*. 2010; 17:51–66. [PubMed: 21218296]
- Kremen WS, Eisen SA, Tsuang MT, Lyons MJ. Is the Wisconsin Card Sorting Test a reliable neurocognitive endophenotype? *American Journal of Medical Genetics (Neuropsychiatric Genetics)*. 2007; 144:403–406.
- Lee T, Mosing MA, Henry JD, Trollor JN, Ames D, Martin NG, Wright MJ, Sachdev PS. OATS Research Team. Genetic influences on four measures of executive functions and their covariation with general cognitive ability: The Older Australian Twins Study. *Behavior Genetics*. 2012; 42:528–538. [PubMed: 22302529]
- Lehto JE, Juujärvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*. 2003; 21(1):59–80.
- Mahone EM, Cirino PT, Cutting LE, Cerrone PM, Hagelthorn KM, Hiemenz JR, Singer HS, Denckla MB. Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology*. 2002; 17:643–662. [PubMed: 14591848]
- Malone SM, Iacono WG. Error rate on the antisaccade task: Heritability and developmental change in performance among preadolescent and late-adolescent female twin youth. *Psychophysiology*. 2002; 39(5):664–673. [PubMed: 12236334]
- 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 the International Neuropsychological Society*. 2010; 16:495–505. [PubMed: 20188014]
- McGue M, Bouchard TJ Jr. Adjustment of twin data for the effects of age and sex. *Behavior Genetics*. 1984; 14:325–343. [PubMed: 6542356]
- Merwood A, Greven C, Price T, Rijdsdijk F, Kuntsi J, McLoughlin G, Asherson P. Different heritabilities but shared etiological influences for parent, teacher and self-ratings of ADHD symptoms: an adolescent twin study. *Psychological Medicine*. 2013; 43(09):1973–1984. [PubMed: 23298428]
- Muthén, LK., Muthén, BO. Mplus User's Guide. 7th. Los Angeles: CA: Muthén & Muthén; 1998–2012.
- Nadder TS, Silberg JL, Eaves LJ, Maes HH, Meyer JM. Genetic effects on ADHD symptomatology in 7-to 13-year-old twins: Results from a telephone survey. *Behavior Genetics*. 1998; 28(2):83–99. [PubMed: 9583234]
- Neale, MC., Boker, SM., Xie, G., Maes, HH. Mx: Statistical Modeling (Version 7th edition). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry; 2006.
- Neiss MB, Stevenson J, Legrand LN, Iacono WG, Sedikides C. Self-Esteem, Negative Emotionality, and Depression as a Common Temperamental Core: A Study of Mid-Adolescent Twin Girls. *Journal of Personality*. 2009; 77(2):327–346. [PubMed: 19192078]

- Nikolas MA, Burt SA. Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: a meta-analysis. *Journal of Abnormal Psychology*. 2010; 119(1):1. [PubMed: 20141238]
- Polderman TJC, Gonso MF, Posthuma D, van Beusterveldt TCEM, Heutink P, Verhulst FC, Boomsma DI. A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *ActaNeurologicaBelgica*. 2006; 106:191–207.
- Power TJ, Werba BE, Watkins MW, Angelucci JG, Eiraldi RB. Patterns of parent-reported homework problems among ADHD-referred and non-referred children. *School Psychology Quarterly*. 2006; 21(1):13.
- Rabbitt P, Lowe C, Shilling V. Frontal tests and models for cognitive ageing. *European Journal of Cognitive Psychology*. 2001; 13(1–2):5–28.
- Rekart KN, Mineka S, Zinbarg RE, Griffith JW. Perceived family environment and symptoms of emotional disorders: The role of perceived control, attributional style, and attachment. *Cognitive Therapy and Research*. 2007; 31(4):419–436.
- Rietveld MJ, Hudziak JJ, Bartels M, Van Beijsterveldt C, Boomsma DI. Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3–12 years. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2003; 117(1):102–113.
- Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*. 2001; 72(5):1394–1408. [PubMed: 11699677]
- Simonoff E, Pickles A, Hervas A, Silberg J, Rutter M, Eaves L. Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*. 1998; 28(04):825–837. [PubMed: 9723139]
- Taylor J. Heritability of Wisconsin Card Sorting Test (WCST) and Stroop Color-Word Test performance in normal individuals: Implications for the search for endophenotypes. *Twin Research and Human Genetics*. 2007; 10:829–834. [PubMed: 18179394]
- Taylor J, Schatschneider C. Genetic influence on literacy constructs in kindergarten and first grade: Evidence from a diverse twin sample. *Behavior Genetics*. 2010; 40:591–602. [PubMed: 20563747]
- Thorell LB, Nyberg L. The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*. 2008; 33(4):536–552. [PubMed: 18568903]
- Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biological Psychiatry*. 2005; 57(11):1336–1346. [PubMed: 15950006]

Figure 1. Confirmatory factor analysis of age-corrected BRIEF scales for the best-fitting model. All path estimates are significant at $p < .001$.

Figure 2.
Cholesky decomposition model estimating rater bias/sibling interaction within twin pair.
Note: A = genetic influences, E = non-shared environmental influences, b = rater bias/
contrast effects.

Table 1

Age distribution of the Sample

Age	<i>n</i>
5	12
6	80
7	82
8	94
9	84
10	116
11	148
12	180
13	194
14	108
15	14
16	2

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Goodness-of-fit Indicators for Confirmatory Factor Analysis Models of the BRIEF

Model	χ^2	df	χ^2 (df)	AIC	CFI	RMSEA
8-Scale (n = 548)						
1-Factor	296.64	20		20531.67	.91	.16
2-Factor	143.76	19	152.88 (1) *	20380.80	.96	.11
3-Factor	124.22	18	19.54 (2) *	20363.25	.97	.10
9-Scale (n = 550)						
1-Factor	417.94	27		22133.55	.89	.16
2-Factor	158.51	26	294.91 (1) *	21876.13	.96	.10
3-Factor	123.03	24	35.48 (3) *	21844.65	.97	.09
Gender Invariance within the BRIEF 9-scale, 3-factor Model (n=550)						
Baseline Model	197.80	54		22255.83	.95	.10
Model 1	208.80	60	11.00 (6)ns	22254.84	.94	.10
Model 2	213.26	66	4.46 (6)ns	22247.30	.94	.09

Note. χ^2 (df) = chi-square and degrees of freedom difference between the 1-factor model and the model shown in the row. Baseline model = fully unconstrained model. Model 1 = factor loadings constrained to be equal across gender groups. Model 2 = factor variances and covariances constrained to be equal across gender groups. Best-fitting models are in bold type.

* $p < .05$.

Table 3
Means and Standard Deviations for Raw BRIEF Scores for the Full Sample and by Gender

	All			Females			Males		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Inhibit	1079	13.10	3.35	578	12.83	3.13	501	13.41	3.58
Shift	1096	11.26	3.06	588	11.25	3.10	508	11.27	3.02
EC	1078	14.17	3.78	576	13.98	3.52	502	14.38	4.05
Initiate	1086	12.03	2.93	583	11.89	2.87	503	12.20	3.01
WM	1087	14.58	4.42	583	14.21	4.10	504	15.01	4.73
Plan/Organize	1086	17.52	5.05	583	17.08	4.78	503	18.04	5.31
Organize	1102	10.78	3.35	589	10.66	3.27	513	10.91	3.44
Monitor	1096	12.27	3.37	585	11.90	3.16	511	12.69	3.55
Self-Monitor	1098	5.79	1.91	589	5.72	1.89	509	5.87	1.93
Task-Monitor	1089	6.56	2.10	582	6.30	2.00	507	6.85	2.18

Note. N refers to number of individuals in the sample. EC = Emotional Control, WM = Working Memory, Organize = Organization of Materials.

Table 4
Summary of Tests Examining Mean and Variance Differences on Age-Corrected BRIEF Scales Across Gender

	Female			Male			<i>t</i>	df	Levene's <i>f</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>			
Inhibit	329	-0.22	3.12	201	0.52	3.30	-2.58*	528	2.25
Shift	336	-0.03	3.12	204	0.16	2.95	-0.70	538	1.38
EC	330	-0.32	3.45	206	0.33	4.16	-1.97*	534	14.10***
Initiate	331	-0.13	2.79	205	0.51	3.06	-2.45*	534	4.03*
WM	329	-0.48	3.96	204	1.11	4.62	-4.25***	531	9.77**
Plan/Organize	331	-0.50	4.69	206	1.27	5.54	-3.95***	535	10.48**
Organize	336	-0.17	3.23	207	0.37	3.43	-1.87	541	0.97
Monitor	332	-0.30	3.27	208	0.83	3.56	-3.77***	538	1.96
Self-Monitor	337	-0.02	1.94	204	0.24	1.94	-1.54	539	0.01
Task-Monitor	334	-0.20	2.08	204	0.52	2.13	-3.87***	536	0.62

Note. N refers to the number of individuals in the sample. One randomly selected twin from each pair was excluded from these analyses. Opposite-sex twins were also excluded from these analyses. EC = Emotional Control, WM = Working Memory, Organize = Organization of Materials.

* $p < .05$.

**

$p < .01$.

$p < .001$.

Table 5

Summary of Tests Examining Mean and Variance Differences on Age-corrected BRIEF Scales Across Zygosity within Gender

		MZ			DZ			<i>t</i>	df	<i>f</i>
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>			
Inhibit	Female	109	-0.39	2.75	128	-0.32	3.17	-0.18	235	2.39
	Male	100	0.20	3.07	98	0.85	3.53	-1.38	196	2.29
Shift	Female	109	0.17	2.98	130	-0.11	3.11	0.72	237	0.08
	Male	101	0.12	2.92	100	0.19	3.02	-0.18	199	0.15
Emotional Control	Female	107	-0.44	2.97	128	-0.10	3.57	-0.78	233	4.15*
	Male	102	-0.03	4.23	101	0.59	4.05	-1.07	201	0.12
Initiate	Female	108	-0.30	2.75	129	-0.27	2.60	-0.08	235	0.78
	Male	100	0.40	2.99	102	0.64	3.16	-0.55	200	0.47
Working Memory	Female	108	-0.81	3.61	128	-0.87	3.44	0.12	234	0.70
	Male	101	1.06	4.41	100	1.23	4.87	-0.26	199	1.28
Plan/Organize	Female	109	-0.66	4.71	129	-1.32	3.65	1.22	236	7.72*
	Male	102	1.40	5.65	101	1.21	5.48	0.25	201	0.01
Organize Materials	Female	108	-0.25	2.59	131	-0.01	3.49	-0.59	237	11.09***
	Male	102	0.20	3.60	102	0.56	3.29	-0.74	202	0.84
Monitor	Female	109	-0.71	2.88	129	-0.72	2.90	0.02	236	0.41
	Male	102	0.73	3.51	103	0.97	3.63	-0.48	203	0.05
Self-Monitor	Female	111	-0.19	1.67	129	-0.14	1.97	-0.22	238	3.17
	Male	102	0.15	1.93	99	0.35	1.97	-0.76	199	1.63
Task-Monitor	Female	107	-0.61	1.76	129	-0.52	1.80	-0.40	234	0.14
	Male	101	0.54	2.14	101	0.51	2.15	0.09	200	0.00

Note. N refers to the number of individuals in the sample. One randomly selected twin from each pair was excluded from these analyses. Opposite-sex twins were also excluded from these analyses.
Organize Materials = Organization of Materials.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 6

Intraclass Correlations for Same-sex and Opposite-sex Twins.

	Female						Male						Opposite-sex					
	MZ			DZ			MZ			DZ			MZ			DZ		
	<i>n</i>	<i>r</i>		<i>n</i>	<i>r</i>		<i>n</i>	<i>r</i>		<i>n</i>	<i>r</i>		<i>n</i>	<i>r</i>		<i>n</i>	<i>r</i>	
Inhibit	108	0.67**		124	0.19*	4.67***	97	0.69**		95	.40**	2.92**	92	0.09				
Shift	108	0.67**		126	0.46**	2.37*	99	0.63**		99	.49**	1.43	97	0.27**				
Emotional Control	105	0.42**		120	0.46**	-0.37	98	0.63**		95	.45**	1.77	94	0.37**				
Initiate	108	0.65**		125	0.24**	4.01***	96	0.69**		98	.50**	2.06*	94	0.20				
Working Memory	108	0.68**		125	0.12	5.36***	99	0.73**		96	.37**	3.74***	92	0.02				
Plan/Organize	109	0.68**		126	0.28**	4.11***	98	0.72**		97	.46**	2.84**	91	0.13				
Organize Materials	108	0.58**		128	0.19*	3.57***	101	0.74**		101	.38**	3.88***	97	0.33**				
Monitor	109	0.62**		125	0.36**	2.64**	99	0.69**		102	.39**	3.07**	94	0.20				
Self-Monitor	109	0.68**		127	0.36**	3.94***	102	0.73**		97	.43**	3.28**	97	0.28**				
Task- Monitor	106	0.41**		123	0.04	2.96**	97	0.62**		99	.24*	3.34***	97	0.07				

Note. N refers to the number of twin pairs. Correlations were tested for significance from zero and, within gender, for difference between MZ and DZ twins using a *z*-test. Organize Materials = Organization of Materials.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 7

Summary of Univariate Biometric Models for BRIEF Scales.

	-2LL	df	AIC	a	c	d	e	b	Gender $\chi^2(df)$	$\chi^2(n)$ vs. Model
Inhibit – Girls										
1 ADE-B	6614.56	1279	4055.66	.85 [.00–.90]	--	.00 [.00–.82]	.15* [.10–.26]	-.20* [-.27–.05]	17.04 (3)	--
2 ADE	6624.19	1281	4062.19	.00 [.00–.26]	--	.73* [.45–.80]	.27* [.20–.37]	--	12.96 (3)	9.63 (1)
3 AE-B	6614.56	128.1	4052.56	.85* [.78–.90]	--	--	.15* [.10–.22]	-.20* [.27–.13]	7.4 (2)	.00 (1)
4 AE	6640.76	128.3	4074.76	.67* [.53–.76]	--	--	.33* [.24–.47]	--	8.07 (2)	26.20 (3)
Inhibit – Boys	--	--	--	--	--	--	--	--	--	--
1 ADE-B	6614.56	127.9	4055.66	.72 [.00–.82]	--	.00 [.00–.68]	.28* [.18–.50]	-.07 [-.16–.09]	--	--
2 ADE	6624.19	128.1	4062.19	.40 [.00–.71]	--	.25 [.00–.72]	.35* [.26–.48]	--	--	--
3 AE-B	6614.56	128.1	4052.56	.72* [.54–.82]	--	--	.28* [.18–.46]	-.07 [-.16–.04]	--	--
4 AE	6640.76	128.3	4074.76	.64* [.51–.73]	--	--	.36* [.27–.49]	--	--	--
Shift – Girls/Boys	--	--	--	--	--	--	--	--	--	--
1 ACE	5493.31	110.8	3277.31	.46* [.23–.67]	.15 [.00–.33]	--	.39* [.32–.47]	--	3.69 (3)	--
2 AE	5495.43	110.9	3277.42	.62* [.55–.69]	--	--	.38* [.31–.45]	--	--	2.12 (1)
3 CE	5508.48	110.9	3290.48	--	.48* [.41–.54]	--	.52* [.46–.59]	--	--	15.17 (1)
Emotional Control – Girls	--	--	--	--	--	--	--	--	--	--
1 ACE	5790.40	107.9	3652.40	.00 [.00–.00]	.37* [.14–.47]	--	.63* [.51–.73]	--	--	--
2 AE	5840.56	107.1	3698.56	.00 [.00–.00]	--	--	1.0* [.10–.1.0]	--	--	50.16 (1)
3 CE	5790.99	107.1	3648.99	--	.37* [.27.47]	--	.63* [.53–.73]	--	--	.59 (1)
Emotional Control– Boys	--	--	--	--	--	--	--	--	--	--
1 ACE	5790.40	107.9	3652.40	.12 [.00–.45]	.46* [.16–.61]	--	.43* [.33–.55]	--	--	--
2 AE	5840.56	107.1	3698.56	.60* [.50–.81]	--	--	.40* [.31–.78]	--	--	--
3 CE	5790.99	107.1	3648.99	--	.54* [.44.62]	--	.46* [.38–.56]	--	--	--
Initiate – Girls/Boys	--	--	--	--	--	--	--	--	--	--

	-2LL	df	AIC	a	c	d	e	b	Gender $\chi^2(df)$	$\chi^2(n)$ vs. Model
1 ADE- B	6335.23	129 0	3755.23	.70 [.00-.78]	--	.00 [.00-.67]	.29* [.22-.45]	-.02 [-.10-.14]	2.74 (3)	--
2 ADE	6335.38	129 1	3753.38	.65* [.16-.75]	--	.04 [.00-.51]	.31* [.25-.39]	--	2.64 (3)	.15 (1)
3 AE-B	6335 23	129 1	3753 23	.70* [.59-.78]	--	--	.30* [.22-.41]	-.02 [-.10-.07]	2.72 (2)	.00 (1)
4 AE	6335.41	129 2	3751.41	.69* [.61-.76]	--	--	.31* [.24-.39]	--	2.75 (2)	.18 (2)
WM - Girls	--	--	--	--	--	--	--	--	--	--
1 ADE- B	7402.63	128 8	4826.63	.64 [.00-.89]	--	.18 [.00-.13]	.18* [.11-.33]	-.16 [-.26-.00]	10.34 (3)	--
2 ADE	7408.47	129 0	4828.47	.00 [.00-.27]	--	.71* [.43-.79]	.29* [.21-.38]	--	8.44 (3)	5.84 (1)
3 AE-B	7402 63	129 0	4822 63	.84* [.76-.89]	--	--	.16* [.11-.24]	-.18* [.26-.11]	10.34 (2)	.00 (1)
4 AE	7424.24	129 2	4840.23	.66* [.53-.76]	--	--	.33* [.24-.47]	--	6.33 (2)	21.61 (3)
WM - Boys	--	--	--	--	--	--	--	--	--	--
1 ADE- B	7402.63	128 8	4826.63	.78 [.00-.86]	--	.00 [.00-.78]	.22* [.14-.43]	-.08 [-.17-.10]	--	--
2 ADE	7408.47	129 0	4828.47	.33 [.00-.76]	--	.39 [.00-.78]	.28* [.21-.38]	--	--	--
3 AE-B	7402 63	129 0	4822 63	.78* [.65-.86]	--	--	.22* [.14-.35]	-.08* [.17-.02]	--	--
4 AE	7424.24	129 2	4840.23	.71* [.60-.78]	--	--	.29* [.22-.40]	--	--	--
Plan - Girls	--	--	--	--	--	--	--	--	--	--
1 ADE- B	7623.73	128 1	5061.73	.00 [.00-.78]	--	.63 [.00-.74]	.37* [.22-.53]	.04 [-.11-.12]	9.33 (3)	--
2 ADE	7624.67	128 3	5058.67	.19 [.00-.70]	--	.47 [.00-.74]	.33* [.26-.43]	--	12.51 (3)	.94 (1)
3 AE-B	7625.62	128 3	5059.62	.71* [.56-.81]	--	--	.30* [.19-.50]	-.05 [-.14-.04]	8.29 (2)	1.92 (1)
4 AE	7627 19	128 5	5057 18	.65* [.55-.73]	--	--	.35* [.27-.45]	--	10.06 (2)	1.57 (3)
Plan - Boys	--	--	--	--	--	--	--	--	--	--
1 ADE- B	7623.73	128 1	5061.73	.69 [.00-.81]	--	.00 [.00-.71]	.30* [.19-.56]	.04 [-.07-.20]	--	--
2 ADE	7624.67	128 3	5058.67	.73* [.35-.80]	--	.00 [.00-.39]	.27* [.20-.35]	--	--	--
3 AE-B	7625.62	128 3	5059.62	.70* [.50-.81]	--	--	.30* [.19-.50]	.03 [-.07-.15]	--	--
4 AE	7627 19	128 5	5057 18	.74* [.65-.80]	--	--	.26* [.20-.35]	--	--	--
Organize - Girls/Boys	--	--	--	--	--	--	--	--	--	--

	-2LL	df	AIC	a	c	d	e	b	Gender $\chi^2(df)$	$\chi^2(n)$ vs. Model
1 ADE- B	6802.47	131 2	4178.47	.75 [.00-.81]	--	.00 [.00-.73]	.25* [.19-.42]	-.06 [-.12-.12]	2.15 (3)	--
2 ADE	6803.17	131 4	4175.17	.32 [.00-.70]	--	.38 [.00-.74]	.30* [.25-.37]	--	1.13 (3)	.70 (1)
3 AE-B	6801.34	131 3	4175.34	.75* [.66-.81]	--	--	.25* [.19-.34]	-.04 [-.12 .04]	.80 (3)	1.13 (1)
4 AE	6806.86	131 5	4176.86	.68* [.61-.74]	--	--	.32* [.26-.39]	--	.17 (3)	5.52 (3)
Monitor - Girls	--	--	--	--	--	--	--	--	--	--
1 ACE	5687.48	110 4	3479.48	.64* [.43-.72]	.00[.00-.00]	--	.36* [.28-.48]	--	--	--
2 AE	5687.48	110 6	3475.48	.64* [.52-.72]	--	--	.36* [.28-.48]	--	--	.00 (1)
3 CE	5840.52	110 6	3628.52	--	.00[.00-.00]	--	1.0* [1.0-1.0]	--	--	153.04 (1)
Monitor - Boys	--	--	--	--	--	--	--	--	--	--
1 ACE	5687.48	110 4	3479.48	.67* [.41-.76]	.00[.00-.00]	--	.31* [.24-.42]	--	--	--
2 AE	5687.48	110 6	3475.48	.67* [.57-.76]	--	--	.31* [.24-.42]	--	--	--
3 CE	5840.52	110 6	3628.52	--	.00[.00-.00]	--	1.0* [1.0-1.0]	--	--	--
SMonitor - Girls/Boys	--	--	--	--	--	--	--	--	--	--
1 ACE	4458.38	110 8	2242.38	.68* [.57-.74]	.00[.00-.00]	--	.32* [.26-.40]	--	.19 (3)	--
2 AE	4458.38	110 9	2240.38	.68* [.60-.74]	--	--	.32* [.26-.40]	--	--	.00 (1)
3 CE	4606.66	110 9	2388.66	--	.00[.00-.00]	--	1.0* [1.0-1.0]	--	--	148.28 (1)
TMonitor - Girls	--	--	--	--	--	--	--	--	--	--
1 ADE- B	5599.57	129 7	3005.57	.74 [.00-.82]	--	.00 [.00-.73]	.26* [.18-.47]	-.22* [-.30-.06]	10.23 (3)	--
2 ADE	5609.34	129 9	3011.34	.00 [.00-.22]	--	.47* [.20-.60]	.53* [.40-.71]	--	9.60 (3)	9.77 (1)
3 AE-B	5599.65	129 9	3001.65	.74* [.59-.82]	--	--	.26* [.18-.41]	-.22* [-.30 .13]	12.15 (2)	.08 (1)
4 AE	5618.24	130 1	3016.24	.34* [.16-.51]	--	--	.66* [.49-.84]	--	10.89 (3)	18.59 (3)
TMonitor - Boys	--	--	--	--	--	--	--	--	--	--
1 ADE- B	5599.57	129 7	3005.57	.28 [.00-.78]	--	.31 [.00-.69]	.41* [.22-.69]	-.01[-.17-.12]	--	--
2 ADE	5609.34	129 9	3011.34	.20 [.00-.64]	--	.38 [.00-.68]	.42* [.32-.56]	--	--	--
3 AE-B	5599.65	129 9	3001.65	.65* [.41-.78]	--	--	.35* [.22-.59]	-.06 [-.17 .06]	--	--

	-2LL	df	AIC	a	c	d	e	b	Gender $\chi^2(df)$	$\chi^2(n)$ vs. Model
4 AE	5618.24	130 1	3016.24	.55* [.42-.66]	--	--	.45* [.34-.58]	--	--	--

Note. Standardized variance estimates are shown with 95% confidence intervals. $\chi^2(df)$ = Chi-square and degree of freedom difference between the gender unconstrained and gender constrained models within a scale. $\chi^2(n)$ vs. Model = Chi-square difference value between current model and parent model (n= number of the parent model). Number of the parent model is designated in the parentheses. S-Monitor = Self-Monitor; Organize = Organization of Materials; WM = Working Memory. Estimates designated with asterisks

* are significant at $p < .05$ based on confidence interval not including zero. The best fitting model for each scale is presented in bold.