Dimensional Structure of the BRIEF2 and its Relations with ADHD Symptoms and Task Performance on Executive Functions in Chinese Children

Kathy Kar-man Shum

Que Zheng

Guinevere Sophia Chak

Kathy Tsam-ling Kei

Cecilia Wing-chi Lam

Iris Kit-yee Lam

Cythia Sin Wah Lok

Jacqueline Wai-yan Tang

The University of Hong Kong

Corresponding Author:

Kathy Kar-man Shum (kkmshum@hku.hk)

Department of Psychology, The University of Hong Kong, Pokfulam Road, Hong Kong

Dimensional Structure of the BRIEF2 and its Relations with ADHD Symptoms and Task Performance on Executive Functions in Chinese Children

Abstract

This study examined the dimensional structure of the Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2) in a Chinese sample of children with attention-deficit/hyperactivity disorder (ADHD)-related concerns and the correlations of the BRIEF2 with the children's ADHD symptoms and their performance on executive function (EF) tasks. Participants were 339 Chinese children aged 6-15 (M = 9.18 years, SD = 2.33; boys: 78.2%) recruited from 35 schools in Hong Kong. The results from confirmatory factor analyses revealed the best fit for a three-factor nine-scale model compared to a two-factor or single-factor model. Significant correlations were found between the BRIEF2 parent and teacher forms for the Behavioral Regulation Index and Cognitive Regulation Index, but not for the Emotion Regulation Index. Associations between performance on an EF task and the rating of the corresponding subscale on the BRIEF2 purportedly measuring the same EF construct were not consistently observed. Lastly, the BRIEF2 showed good convergent validity with the ratings of ADHD symptoms on the Swanson, Nolan, and Pelham Rating Scale Version IV (SNAP-IV). This study provided plausibly the first evidence on the dimensional structure of the BRIEF2 Parent and Teacher Forms in an Asian sample and confirmed the factorial validity of the Chinese version of the BRIEF2.

Dimensional Structure of the BRIEF2 and its Relations with ADHD Symptoms and Task Performance on Executive Functions in Chinese Children

Introduction

Executive functions (EF) are a set of top-down and inter-related mental processes that enable us to filter distractions, prioritize tasks, set goals, and control impulses. They belong to a multifactorial construct encompassing skills that are developed over time and essential in daily life (Jacobson et al., 2019). These cognitive processes include inhibitory control, attentional control, working memory, planning, organization, self-regulation, initiation and monitoring of tasks, and cognitive flexibility (Barkley, 2012; Chan et al., 2008; Diamond, 2013).

Executive dysfunction is a key feature of many developmental disorders, such as attention-deficit/hyperactivity disorder (ADHD). Neuropsychological models traditionally emphasize the role of executive dysfunction in ADHD pathophysiology, which gives rise to symptoms of inattention, hyperactivity and impulsivity (Barkley, 1997; Bayliss & Roodenrys, 2000). A meta-analysis of 83 empirical studies using performance-based measures (Willcutt et al., 2005) indicated significant impairments in EF among children with ADHD when compared to their peers with typical development, especially in the domains of inhibitory control, planning, vigilance and working memory. Nonetheless, it is noteworthy that fewer than half of the ADHD samples (*N*=3734) in the meta-analysis exhibited impairment in any specific EF tasks (Willcutt et al., 2005). On one hand, this finding suggested that EF weaknesses might not be present in all cases of ADHD. On the other hand, it also called attention to the reliability of the performance tasks used in measuring EF (Willcutt et al., 2005).

Apart from performance-based measures, behavioral ratings reported by parents and teachers are widely used to evaluate children's executive behaviors. This type of measures highlights the importance of the ecological component in the assessment of EF, as they reflect children's actual application of EF skills in daily tasks (Jacobson et al., 2019; Pino Muñoz & Arán Filippetti, 2019).

The Behavior Rating Inventory of Executive Function as a Measure of EF

The Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000) is an ecological assessment tool based on the report of parents and teachers to evaluate the behavioral, emotional and cognitive competence in executive functioning for children and adolescents aged 5-18 (Davidson et al., 2016; Halvorsen et al., 2019). Empirical evidence has supported the usefulness and validity of the BRIEF in assessing EF skills observed by parents and teachers in different functional settings for children with neuropsychological developmental disorders such as ADHD (Mahone et al., 2002) and autism spectrum disorder (ASD; Gilotty et al., 2002).

The original version of the BRIEF (Gioia et al., 2000) contains 86 items categorized into eight scales and two indexes: the Emotional Control, Shift, and Inhibit scales are grouped under the Behavior Regulation Index (BRI); and five other scales, namely Plan/Organize, Working Memory, Initiate, Organization of Materials, and Monitor, are grouped under the Metacognition Index (MI). The BRI and MI together constitute the Global Executive Composite (GEC). The dimensional structure of the original BRIEF has been validated as a two-factor, eight-scale model in a number of studies (e.g., Egeland & Fallmyr, 2010; Halvorsen et al., 2019; Jacobson et al., 2016; Lyons Usher et al., 2016; McCandless & O'Laughlin, 2007). However, other studies expressed uncertainty towards the two-factor model for the BRIEF. For instance, a confirmatory factor analysis (CFA) of a sample of 281 children and adolescents with different neurodevelopmental disorders found that a three-factor model with nine scales showed the most appropriate fit for the BRIEF parent form (Halvorsen et al., 2019). Two other CFAs of the BRIEF teacher form also found evidence supporting a model with three factors, rather than two (Egeland & Fallmyr, 2010; Peters et al., 2012).

Development of the BRIEF2 and its Factorial Validity

Based on evidence supporting a three-factor model, a revised version—the Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2)—was published in 2015 (Gioia et al., 2015). It contains 63 items subsumed under three factors (three indexes) and nine scales. The Inhibit and Self-Monitor scales are categorized under the Behavior Regulation Index (BRI), which measures a child's ability to regulate and monitor his or her own behavior. The Shift and Emotional Control scales are categorized under the Emotional Regulation Index (ERI), which assesses a child's ability to regulate and adapt emotional responses. Five scales including Initiate, Working Memory, Plan/Organize, Task-Monitor, and Organization of Materials scales are grouped under the Cognitive Regulation Index (CRI), which measures a child's ability to control and manage cognitive processes and to solve problems (Gioia et al., 2015).

So far, very few studies have validated and confirmed the dimensional structure of the BRIEF2. Jiménez and Lucas-Molina (2019) conducted CFA to evaluate the factorial structure of the BRIEF2 parent form across genders in a sample of school-age children in the Dominican Republic and found that the three-factor nine-scale model was a good fit. Their findings were corroborated by results from another CFA study, which also reported a good fit for the three-factor nine-scale model for both the parent and teacher forms among typically developing children aged 8 to 12 in Chile (Pino Muñoz & Arán Filippetti, 2019). Jacobson et al. (2019) examined the factorial validity of the BRIEF2 parent form using exploratory factor analysis

(EFA) and CFA on a clinical sample of 5212 children and adolescents aged 5-18 in the United States and observed that a model of three factors with at most six scales was a better fit. Despite slight differences in the factorial structure of the BRIEF2 found between clinical (Jacobson et al., 2019) and typically developing populations (Jiménez & Lucas-Molina, 2019; Pino Muñoz & Arán Filippetti, 2019), the findings were rather similar across these studies. Nevertheless, with only a limited number of studies that have investigated the dimensional structure of the BRIEF2, it awaits further investigation before more conclusions can be drawn.

Moreover, as prior studies on the structure of both the BRIEF and the BRIEF2 mostly involved samples from Europe (e.g., Norway: Egeland & Fallmyr, 2010; France: Fournet et al., 2015; Netherlands: Huizinga & Smidts, 2011) and North and South America (e.g., United States: Jacobson et al., 2019; Mahone et al., 2002; Peters et al., 2012; Dominican Republic: Jiménez & Lucas-Molina, 2019; Chile: Pino Muñoz & Arán Filippetti, 2019), less is known about the validity of the dimensional structure of the BRIEF/BRIEF2 when used in other regions around the world. To the best of our knowledge, no study so far has evaluated the dimensional structure of the BRIEF2 among non-Western populations. It remains largely unknown whether the factorial model of the BRIEF2, which was developed and normed based on samples in the United States, is also applicable to non-Western populations. There is a need for more studies that assess the factorial validity of the BRIEF2 parent and teacher forms in different cultural contexts in order to understand both the applicability of the instrument cross-culturally and the cultural differences in everyday executive functions. The current study aimed to address this research gap by examining the dimensional structure of both the parent and teacher forms of the BRIEF2 among Chinese school-aged children.

Do Ratings on the BRIEF2 Parent and Teacher Forms Correlate?

The study by Pino Muñoz and Arán Filippetti (2019)—the only published study that investigated the relation between the BRIEF2 parent and teacher ratings—reported low to moderate associations between the parent and teacher reports, with the corresponding CRI of the two forms showing the highest correlation (r = .65), followed by the BRI (r = .50), and lastly the ERI showing the lowest but still significant correlation (r = .30). Pino Muñoz and Arán Filippetti (2019) stated that the significant correlations between the parent and teacher forms of the BRIEF2 provided support for the equivalence of the forms in measuring similar constructs of EF. Nevertheless, the discrepancies in the strengths of correlations between indexes in the two forms indicated a heterogeneity of the informants' ratings across scales.

Some scholars have proposed that EF can be viewed as either "cold" or "hot" (Zelazo et al., 2004). Cold EF refers to those EF constructs involved in logical reasoning and analysis, as well as the conscious control of thoughts and behaviors with little emotional salience (Rubia, 2011). On the other hand, hot EF refers to the affective cognitive abilities that are involved in affective decision making (Zelazo & Carlson, 2012; Zelazo & Muller, 2002). As such, we hypothesized that the CRI of the BRIEF2 might relate more to cold EF while the ERI is an indicator of hot EF. As for the BRI, it comprises the Inhibit and Self-Monitor scales which measure children's impulsivity, and therefore more likely reflects hot EF. The stronger correlation of the CRI between parents' and teachers' ratings might imply that cold EF skills are perceived more consistently by informants across varied daily settings. The weaker correlations of the BRI and ERI, by contrast, might indicate that the informants' perceptions of hot EF skills differ more across settings.

Apart from settings, variance in ratings may also be due to varying degrees of sensitivity towards children's typical development in the EF domain. It has been reported in some studies that teachers are better than parents in discerning children's difficulties in EF, because of their training and experience with age-appropriate behaviours (Bausela-Herreras, 2018; Mares et al., 2007). Nevertheless, existing research on the relations between the parent and teacher forms of the BRIEF2 remains scarce. The extent to which the parent- and teacher-reports of the BRIEF2 concurrently reflect the actual executive functioning of the child deserves further exploration.

Do the BRIEF2 Ratings Correlate with EF Task Performance?

Performance-based tasks that measure EF in children have been widely used in many studies (Barkley, 1997; Elosúa et al., 2017; Lawrence et al., 2004; Sjöwall et al., 2013; Sonuga-Barke et al., 2008). A meta-analytical review indicated that deficits in EF among children with ADHD were observed on performance-based measures of inhibition, vigilance, working memory, and planning (Willcutt et al., 2005). There are a vast variety of standardized performance-based tasks that attempt to measure different domains of EF, such as the Conners' Continuous Performance Test (CPT) which measures inhibition and vigilance (Conners, 2014); the Tower of London test which measures planning (Shallice, 1982; Culbertson & Zillmer, 1998); digit span and picture span tasks that assess working memory (Cohen, 1997); and the Children's Color Trails Test that measures set-shifting (Williams et al., 1995).

Prior research has reported that the performance on EF tasks and behavioral ratings on the BRIEF did not closely correlate with each other. For instance, McAuley and colleagues (2010) found that although both the parent and teacher BRIEF ratings correlated significantly with symptoms of ADHD such as inattentiveness, impulsivity and hyperactivity, the BRIEF ratings were not associated with any EF task performance scores. Similarly, a study of adolescents with ADHD found that the test scores on inhibition, working memory, set-shifting, and planning correlated significantly with the overall BRIEF ratings but did not uniquely associate with the respective BRIEF scale scores (Toplak et al., 2009). A practitioner review of 13 studies involving the BRIEF and performance-based EF tasks reported an overall median correlation of .18 between these two types of measures (Toplak et al., 2013). Based on those 13 studies on the BRIEF included in the review, only 35 (i.e., 19%) of the 182 correlational comparisons between the performance-based measures and the indexes and subscales of the BRIEF reached statistical significance (Toplak et al., 2013).

Toplak et al. (2013) postulated that the weak correlations between the BRIEF ratings and EF task performance might suggest that they are tapping cognitive functioning on two different levels—the algorithmic and the reflective levels. Performance on the algorithmic level can be explained by information processing mechanisms, such as perceptual regulation and working memory, and is typically related to work efficiency (Anderson, 1990). The reflective level of analysis, by contrast, concerns the goals of the person, beliefs related to the goals, and the rational actions taken in alignment with the goals and beliefs (Bratman et al., 1991). It has been suggested that task-based performance may indeed be measuring the efficiency of EF processes at the algorithmic level, while bypassing the domain of rational goal pursuit, as the goals and structure of EF tasks are often provided by the experimenter without the need for discovery or creation on the part of the research participants (Salthouse et al., 2003). By contrast, behavioral ratings of EF based on daily observations provide information about a person's success in rational goal pursuit, and they may be more related to the reflective level of cognitive functioning. Therefore, the results on executive functioning derived from these two types of EF measures might appear different. In line with this argument, Gerst et al. (2017) found weak correlations (rs < .25) between the BRIEF ratings and EF task performance but they were both equally and moderately correlated with academic measures (rs ranging from .29 to .50), suggesting that the two types of measures might be capturing different aspects of EF but both are important in predicting real-world functioning.

Pino Muñoz and Arán Filippetti (2019) examined and reported the relations between EF task-performance and the BRIEF2 ratings. Similar to earlier findings on the BRIEF (Gerst et al., 2017; McAuley et al., 2010; Toplak et al., 2009; Toplak et al., 2013), low correlations (ranging from .19 to .32) were found between task performance and the BRIEF2 ratings on both the teacher and parent forms (Pino Muñoz & Arán Filippetti, 2019). Needless to say, more evidence is required to determine whether the BRIEF2, as an indirect measure of EF via other-reports, produces results that align with the actual direct observation of children's EF performance via task measures, especially among those who have difficulties in executive functioning.

To What Extent do the BRIEF2 Ratings Predict ADHD Symptoms?

Another important issue is whether and to what extent the BRIEF2 associates with symptoms of ADHD, i.e., inattention and/or hyperactivity/impulsivity as described in the diagnostic criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013). A strong correlation between the BRIEF2 and ADHD symptoms would suggest that the instrument has high concurrent validity on identifying children with ADHD.

While many studies in the past have investigated the correlations of the BRIEF ratings and ADHD symptoms, relatively less have examined the associations of the BRIEF ratings with ADHD-subtypes. One such study found that the BRI of the BRIEF could discriminate the ADHD-combined (ADHD-C) subtype—those with significant symptoms of both inattention and hyperactivity/impulsivity—from the ADHD-inattentive (ADHD-I) subtype and non-ADHD groups, whereas the MI could discriminate both the ADHD-C and ADHD-I from the non-ADHD group (McCandless & O'Laughlin, 2007). In another study based on a sample of youths with ADHD and a non-ADHD comparison group, higher BRI and ERI on the BRIEF2 parent form were associated with greater hyperactivity, whereas higher CRI was correlated with greater inattentiveness (Jacobson et al., 2016). However, the correlations of the BRIEF2 teacher form and ADHD symptoms have not been examined thus far.

The Current Study

This study aimed to investigate: 1) the dimensional structure of the Chinese version of the BRIEF2 parent and teacher forms in a sample of children with ADHD-related concerns using CFA; 2) the relations between the ratings on the parent and teacher forms of the BRIEF2 by examining the correlations between the corresponding indexes and subscales on the two measures; 3) whether the BRIEF2 ratings correlate with performance-based measures of EF; and 4) whether the BRIEF2 ratings correlate with ADHD symptoms and to what extent the indexes of the BRIEF2 predict inattention and hyperactivity/impulsivity symptoms that correspond directly to the diagnostic criteria of ADHD.

Method

Participants and Procedures

A total of 339 native Chinese students aged 6 to 15 (mean age = 9.18 years, SD = 2.33; boys: 78.2%) at junior elementary (n = 151; mean age = 7.24 years, SD = 0.61; boys: 80.8%), senior elementary (n = 114; mean age = 9.22 years, SD = 0.59; boys: 73.7%), and junior secondary school grades (n = 74; mean age = 13.09 years, SD = 0.71; boys: 79.7%) were recruited from 35 schools in Hong Kong to participate in the study. All participants were nominated by the schools to join an intervention program that targeted students with ADHD-related concerns to enhance their EF. Written informed consent was obtained from the parents and the children before inclusion in the study. Ethical approval for this research was obtained from the Human Research Ethics Committee at the authors' university.

The demographic characteristics of the participants are presented in Table 1. In our sample, 71.1% of the students had a clinical diagnosis of ADHD while another 7.7% were suspected of having ADHD, as reported by psychiatrists and pediatricians in previous clinical assessment reports. As the participants were initially recruited for an intervention that targeted students with ADHD, students with a diagnosis of ASD were excluded from the study. Hence, despite considerable comorbidity typically reported between ADHD and ASD (Antshel et al., 2016), there were only 2.7% of the students in our sample with suspected ASD. Other confirmed diagnoses of comorbid conditions included intellectual disability (0.3%), oppositional defiant disorder (2.7%), and dyslexia (24.2%). The diagnoses were not mutually exclusive. According to the parent reports, an average of 47.4% of the students across all ages used ADHD medication. An increasing trend was observed in medication use from younger to older children.

Parents and teachers of the participating students were invited to complete a questionnaire on the child's EF and ADHD symptoms. Parents further provided demographic information about the child's medical history and the use of ADHD medication. Data included in the current analyses were obtained from the pretest assessment of the children before entering the training program. All students were assessed on their IQ, working memory, planning, shifting, attention, and inhibition by completing individual tasks administered by qualified educational psychologists and trained research assistants, during regular school hours at the participating schools.

Measures

Behavior Rating Inventory of Executive Function, Second Revision (BRIEF2). Both the BRIEF2 Parent and Teacher Forms (Gioia et al., 2015) consist of 63 items across nine subscales grouped under three indexes: the Behavior Regulation Index (BRI) is computed from the Inhibit and Self-Monitor scales; the Emotion Regulation Index (ERI) from the Shift and Emotional Control scales; and the Cognitive Regulation Index (CRI) comprises the Initiate, Working Memory, Plan/Organize, Task-Monitor, and Organization of Materials scales. The three indexes together form the Global Executive Composite (GEC). Parents and teachers of the children participants were asked to indicate how often the described behaviors generate problems in daily life over the past six months on a 3-point Likert scale ("never", "sometimes", and "often"), with higher scores signifying more pronounced problems encountered. The Chinese version of the BRIEF2 Parent Form was obtained directly from the publisher of the questionnaire (Gioia et al., 2015). The Chinese version of the BRIEF2 Teacher Form was translated and back-translated by our research team with the permission of the publisher (Psychological Assessment Resources, Inc.; PAR), and the version used in our data collection was approved by PAR. Raw scores from the BRIEF2 parent and teacher forms were used in the analyses of this study. The internal consistencies of the scales and indexes for the Parent and Teacher Forms are shown in Table 2.

Swanson, Nolan, and Pelham Rating Scale Version IV (SNAP-IV). The SNAP-IV is a widely used scale that measures the core symptoms of ADHD. The Chinese version of the SNAP-IV Parent and Teacher Forms (Gau et al., 2008a, 2008b) was adopted in the current study. Both forms consist of 18 items—with 9 items each on Inattention (IA) and Hyperactivity/Impulsivity (HI) respectively—that correspond directly to the symptoms described in the diagnostic criteria of ADHD in DSM-5 (American Psychiatric Association, 2013). As the focus of the current study was on the association between the BRIEF2 and ADHD symptoms, the other 8 items on the SNAP-IV that correspond to the diagnostic criteria of oppositional defiant disorder were not included in our analyses.

Parents and teachers rated each item on a 4-point Likert scale to indicate how severe the symptom was (“0” = not at all, “1” = just a little, “2” = quite a bit, and “3” = very much). In this study, we regarded a rating of “2” or above on an item as indicating the presence of that symptom to a notable degree (Gau et al., 2008b). Since one of the diagnostic criteria of ADHD in DSM-5 for children under age 17 requires the display of at least six out of the nine symptoms on either IA or HI, considering both the severity (i.e. the raw scores) and number of symptoms on the scales enabled us to examine how well the BRIEF2 correlated with the diagnostic symptoms of ADHD as informed by the scores on the SNAP-IV.

Wechsler Nonverbal Scale of Ability (WNV). The two-subtest battery of the WNV (Wechsler & Naglieri, 2006) was used to yield the estimated IQ of the participants. The two subtests included Matrices and either Recognition (for children aged 7 or below) or Spatial Span (for those aged 8 or above).

Children Memory Scale (CMS). Two subtests—Numbers and Picture Locations—from CMS (Cohen, 1997) were selected as measures of auditory and visual working memory respectively.

Tower of London Test (TOL). The TOL measured the children’s ability to plan a course of action in a problem-solving task (Shallice, 1982). Score calculation was based on both the execution time and the number of attempts required for successful completion of each problem (Anderson et al., 1996). A higher score denotes better planning and problem-solving skills.

Children's Color Trails Test (CCTT). The CCTT is a paper-and-pencil test designed to assess sustained attention, sequencing, and cognitive flexibility (Llorente et al., 2003). The interference score was computed (Llorente et al., 2003), and a higher interference score indicates more susceptibility to interference and lower cognitive flexibility (Llorente et al., 2009).

Conners' Continuous Performance Test (CPT). The Conners Continuous Performance Test 3rd Edition (CPT 3; Conners, 2014) and Conners Kiddie Continuous Performance Test 2nd Edition (K-CPT 2; Conners, 2015) were conducted respectively for children aged 8 or above, and those below 8, to assess their ability to sustain attention over a period of time and to inhibit automatic responses. The following measures from CPT were used in our analyses: d-prime (*d'*), omission error rate, and commission error rate.

Statistical Analyses

To verify the dimensionality of the Chinese version of the BRIEF2 Parent and Teacher Forms, confirmatory factor analyses were conducted using R (R Core Team, 2019) and the packages of "psych" (Revelle, 2018), "lavaan" (Rosseel, 2012), and "semTools" (Jorgensen et al., 2019) to examine the goodness-of-fit of the a priori models. Model parameters were estimated using the diagonally weighted least squares (DWLS) procedure, which was specifically suited for ordinal data in questionnaires. In addition to the overall χ^2 statistics, other indicators of goodness-of-fit with the following "rule-of-thumb" cutoff criteria for well-fitting models were used: standardized root mean squared residual (SRMSR) ≤ 0.08, root mean squared error of approximation (RMSEA) ≤ 0.06, Tucker–Lewis Index (TLI) and Comparative fit index (CFI) ≥ 0.95 (Hu & Bentler, 1999).

To explore how well the BRIEF2 Parent and Teacher Forms corelated, and how well the EF domains rated on the BRIEF2 correlated with other behavioral ratings and performance-based measures, correlational analyses were employed to examine the associations between: 1) the BRIEF2 Parent and Teacher Forms; 2) the BRIEF2 and performance-based measures of EF; and 3) the BRIEF2 and the SNAP-IV. Hierarchical regression analyses were run to further investigate the unique contribution of the BRIEF2 indexes in predicting IA and HI symptoms on the SNAP-IV after controlling for student's age, gender, and IQ scores. Gender was included as a controlled variable in the analyses as gender differences in the hyperactivity level and rate of externalizing behaviors in ADHD samples have been reported (Gaub & Carlson, 1997).

Results

Dimensional Structure of the BRIEF2 Parent and Teacher Forms

Table 2 shows the means, standard deviations, internal consistencies, and the inter-correlations among the nine subscales for the BRIEF2 Parent (BRIEF-P; above the diagonal) and Teacher Forms (BRIEF-T; below the diagonal). For the parent form, all the nine subscales and index scores were moderately to highly correlated. Correlation coefficients ranged from .38 to .96 (all ps < .001). The internal consistencies across items were good. Cronbach's alpha coefficients for each subscale ranged from .66 to .89, and the Cronbach's alphas for each index (i.e., BRI, ERI, CRI, and GEC) ranged from .86 to .95.

For the teacher form, all the nine subscales and index scores were moderately to highly correlated. Correlation coefficients ranged from .26 to .97 (all ps < .001). The internal consistencies were good to excellent. Cronbach's alpha coefficients for each subscale ranged from .81 to .94, and the Cronbach's alphas for each index ranged from .92 to .97.

Confirmatory factor analyses (CFA) were conducted separately for the BRIEF2 Parent and Teacher Forms to test the model fit of the 9-factor model proposed in earlier studies (Gioia et al., 2015). Table 3 summarizes the relevant statistics of the model fits for the BRIEF-P and the

BRIEF-T. The 9-factor model showed a very good fit for both the BRIEF-P and the BRIEF-T, suggesting that the nine-scale model was applicable to the Chinese version of the BRIEF2 for both the parent and teacher forms.

CFA was further conducted to compare the following models: 1) a three-factor model (behavior regulation [BR], emotion regulation [ER], and cognitive regulation [CR]) proposed recently in the development of the BRIEF2 (Jiménez & Lucas-Molina, 2019; Pino Muñoz & Arán Filippetti, 2019); 2) two-factor models, with two of the three indexes (BRI, ERI or CRI) combined and loaded as one factor in each model (Jiménez & Lucas-Monlina, 2019); and 3) a simple one-factor model, with all subscales loaded on one single factor. The mean raw scores for each of the nine subscales were entered as measured variables in each tested model.

The results of the CFA for both the BRIEF2 Parent and Teacher Forms were similar, indicating the best model fit for the three-factor model compared to the two- and one-factor models (Table 3). The model fit indexes of the one-factor model showed a poorer fit of this model than the other tested models. Nonetheless, it should be noted that the one-factor model for the BRIEF-P was still considered well-fitted according to the "rule-of-thumb" cutoff criteria stated earlier. Figure 1 shows the factor loadings of the nine subscales on the three factors (BRI, ERI, and CRI) for the BRIEF2 Parent and Teacher Forms.

Correlations between Ratings on the BRIEF2 Parent and Teacher Forms

Pearson correlations between 1) the corresponding indexes and 2) the corresponding subscales of the BRIEF-P and the BRIEF-T were computed (Table 4). Both the BRI (r = .14, p = .009) and CRI (r = .13, p = .02) showed significant but weak correlations between the BRIEF-P and the BRIEF-T. No significant associations were found between the two forms on the ERI (r = .04, p = .51) and GEC (r = .05, p = .39). At the subscale level, the Inhibit subscale, but not the

Self-Monitor subscale, of BRI on the BRIEF-P correlated significantly with its corresponding subscale on the BRIEF-T. For the five CRI subscales, three of them (i.e., Initiate, Task-Monitor, and Organization of Materials) on the BRIEF-P each significantly correlated with its corresponding subscale on the BRIEF-T. No significant correlations were found between the Parent and Teacher Forms on the ERI subscales (i.e., Shift and Emotional Control).

Correlations between the BRIEF2 Ratings and Performance-Based Measures of EF

Table 5 presents the Pearson correlation coefficients between the BRIEF2 indexes and direct cognitive measures of children's IQ and EF. The BRIEF-P and the BRIEF-T displayed different patterns of correlations with the assessment measures of EF. Among the various EF tasks, the BRIEF-P correlated more with the working memory tasks—namely Numbers and Picture Locations—on the Children Memory Scale (correlations with the GEC ranging from -.12 to -.20, *p*s < .05), while the BRIEF-T ratings correlated more strongly with performance on the CCTT and the CPT (correlations with the GEC ranging from .13 to .20, *p*s < .05).

Importantly, significant associations were not consistently revealed between performance on an EF task and the rating of the corresponding subscale on the BRIEF2 purportedly measuring the same EF domain. In particular, performance on Picture Locations which assessed working memory significantly correlated with the Working Memory scales on both the BRIEF-P (r = -.17, p = .002) and the BRIEF-T (r = -.14, p = .01), but Numbers only correlated with the Working Memory subscale on the BRIEF-T (r = -.14, p = .009) and not the BRIEF-P (r = -.10, p = .08). Moreover, scores on the TOL did not associate significantly with ratings of the Plan subscale on either the parent (r = .00, p = .94) or the teacher form (r = -.10, p = .07), nor did the interference score of the CCTT correlate with the Shift subscale on either the BRIEF-P (r = .00, p = .94) or the BRIEF-T (r = .10, p = .07). Similarly, the commission errors of CPT, presumably indicating inhibitory control, showed no significant correlations with the Inhibit subscales on both the BRIEF-P ($r = .05$, $p = .40$) and the BRIEF-T ($r = .06$, $p = .24$).

It is worth mentioning that the IQ score for the older participants in this study was estimated partly based on the Spatial Span subtest, which is conceptually similar to the Picture Locations subtest for measuring working memory. Nonetheless, IQ score did not correlate significantly with the Working Memory subscale on either the BRIEF-P ($r = -.06$, $p = .28$) or the BRIEF-T ($r = -.08$, $p = .16$).

Correlations between the BRIEF2 Ratings and ADHD Symptoms on the SNAP-IV

Correlations between the BRIEF2 and the SNAP-IV parent and teacher ratings were examined to see how well the various domains of EF associated with the severity (indicated by the raw scores) and the number of ADHD symptoms measured by the SNAP-IV (Table 6).

As expected, stronger correlations were observed between the BRIEF2 and the SNAP-IV rated by the same informants than across different informants (i.e., parents vs. teachers). For the BRIEF-P, all the indexes showed moderate correlations with both the IA and HI symptoms on the SNAP-IV Parent Rating (SNAP-P). Among the three indexes, the CRI and BRI showed the strongest correlations with IA and HI respectively. Similar results were obtained between the BRIEF-T and the SNAP-IV Teacher Form (SNAP-T). Specifically, the CRI showed strong associations with IA symptoms on the SNAP-T, and the BRI showed strong correlations with HI symptoms.

Across informants, much weaker correlations were observed. Nevertheless, the CRI on the BRIEF-T correlated significantly with symptoms of IA on the SNAP-P, while the BRI correlated significantly with HI. Analogous correlational patterns were found between the BRIEF-P and the SNAP-T. Lastly, significant correlations were indicated across informants between the SNAP-P and the SNAP-T for both IA (severity: $r = .16$, $p = .003$; no. of symptoms: $r = .20$, $p < .001$) and HI symptoms (severity: $r = .33$, $p < .001$; no. of symptoms: $r = .33$, $p < .001$). These correlations, especially for HI, appeared slightly more robust than the correlations of indexes (ranging from .04 to .14) between the BRIEF-P and the BRIEF-T.

Hierarchical Regressions Using the BRIEF2 Index Scores to Predict Inattention and Hyperactivity/Impulsivity Symptoms on the SNAP-IV

We performed a series of hierarchical regressions to evaluate the relative contribution of the BRIEF2 domains to the severity and the number of symptoms of IA and HI on the SNAP-IV Rating Scales. The results of the regression analyses for the SNAP-P and the SNAP-T are presented in Tables 7 and 8 respectively. In each regression model, students' age, gender, and IQ scores were entered as controlled variables in the first step. In the second step of the regressions, the three composite indexes (BRI, ERI, and CRI) of either the BRIEF-P or the BRIEF-T were entered simultaneously, and all 6 indexes from both the BRIEF-P and the BRIEF-T were included in the final step of the analyses.

In the prediction of IA and HI symptoms on the SNAP-P (Table 7), the three indexes of the BRIEF-P together contributed over 25% of variance in each outcome measure beyond the controlled variables ($\Delta R^2 \geq .25$; ps $\leq .001$), whereas the indexes of the BRIEF-T only accounted for 5-6% of the variance in the outcome measures (Step 2; Table 7). Among the three indexes in Step 2 of the regression equations, the CRI of either the BRIEF-P or the BRIEF-T was the strongest predictor of IA symptoms and the BRI was the strongest predictor of HI symptoms on the SNAP-P. When all indexes were entered simultaneously in Step 3 of the analyses, the BRIEF-P CRI was the strongest predictor of IA (βs $\geq .40$; ps $\leq .001$), followed by the BRIEF-T CRI (βs $\geq .14$; ps $\leq .05$), over and above all other variables. Similarly, the BRIEF-P BRI was the strongest predictor of HI in Step 3 of the analyses (βs ≥ .46; ps ≤ .001), while additional unique variance was contributed by the BRIEF-T BRI (βs ≥ .26; ps ≤ .001).

For the prediction of the SNAP-T ratings (Table 8), the indexes of the BRIEF-T together contributed nearly half of the total variance of IA and HI symptoms in Step 2 after controlling for the demographic variables ($\Delta R^2 \geq .47$; ps ≤ .001). The contribution of the BRIEF-P was significant in predicting HI on the SNAP-T ($\Delta R^2 = .06$; ps ≤ .001), but not significant in the prediction of IA. In the final step of the regression when all variables were entered in the model, the BRIEF-T CRI was the only index significant in predicting both the severity and number of IA symptoms on the SNAP-T (βs ≥ .62; ps ≤ .001). The BRIEF-T CRI was also found to be a significant negative predictor of HI (βs ≥ -.20; ps ≤ .001), and the BRIEF-T BRI remained the strongest predictor above all other variables in the prediction of HI on the SNAP-T (βs ≥ .77; ps ≤ .001).

Discussion

In the present study, we examined the following: 1) the factorial validity of the three-factor nine-scale model for the Chinese version of the BRIEF2 parent and teacher forms in a sample of children with ADHD-related concerns; 2) the correlations between the corresponding indexes and subscales on the parent and teacher forms of the BRIEF2; 3) the correlations between the BRIEF2 ratings and performance-based measures of EF; and 4) the correlations between the BRIEF2 ratings and ADHD symptoms and the extent to which indexes of the BRIEF2 predicted inattention and hyperactivity/impulsivity symptoms measured on the SNAP-IV.

Firstly, we examined the dimensional structure of the Chinese version of the BRIEF2 parent and teacher forms among a sample of children aged 6-15. The Chinese version of the

BRIEF2 Teacher Form was translated by our research team and approved by the publisher (PAR). Hence, this study was the first to explore the validity of its factorial structure and provided support for the three-factor nine-subscale model in contrast to a two-factor or single-factor model based on confirmatory factor analyses. The 3-factor model showed the best model fit for both the BRIEF-P and the BRIEF-T in contrast to other tested models. Our results corroborate the findings of earlier studies (Jiménez & Lucas-Molina, 2019; Pino Muñoz & Arán Filippetti, 2019), and provide plausibly the first evidence on the dimensional structure of the BRIEF2 Parent and Teacher Forms among an Asian sample outside Europe and America. The results here confirm the factorial validity of the BRIEF2 in a Chinese sample of school-aged children with ADHD-related concerns, thereby providing evidence to support the applicability of the BRIEF2 in different cultural contexts. The finding of a similar factor structure of the BRIEF2 across cultures may perhaps suggest that the EF construct is more biologically determined and less susceptible to cultural influences (Boivin & Giordani, 2009).

Secondly, we examined the relations between the ratings on the parent and teacher forms of the BRIEF2 to test how well the corresponding indexes and subscales on the two measures correlated with each other. Pino Muñoz and Arán Filippetti (2019) reported low to moderate associations between the indexes of the parent and teacher reports, with *r*s ranging from .30 (for the ERI) to .65 (for the CRI). In our study, weak significant correlations were found between the BRIEF-P and the BRIEF-T for the BRI (r = .14) and CRI (r = .13), but not for the ERI (r = .04). It is noteworthy that data were collected on typically developing school-aged children in Pino Muñoz and Arán Filippetti (2019), whereas the current sample included mainly children with ADHD. The BRIEF2 manual (Gioia et al., 2015) reported weaker correlations between parent and teacher ratings in mixed clinical samples (ranging from .30 to .46) than in typically developing children (ranging from .50 to .72), suggesting that the extent of interrater agreement might vary across clinical and non-clinical samples. Associations between parent and teacher ratings in the present study were even lower than those documented by Gioia et al. (2015), perhaps due to the sampling of individuals with mainly a single diagnosis of ADHD in contrast to the mixed clinical samples based on which correlations were reported in the BRIEF2 manual.

Similar to the findings of Pino Muñoz and Arán Filippetti (2019), the ERI of the two BRIEF2 forms showed the weakest correlation among the three indexes. Assuming that the ERI relates more to hot EF, our results are in line with the proposition that hot EF skills are perceived less consistently by informants across settings. The behavioral manifestations of emotion regulation may vary across settings to a greater extent than behaviors of cognitive and behavioral regulation. The discrepancies between the parent and teacher ratings might have been accentuated by the classroom context in Hong Kong, where class size is typically about 30-40 students and students are offered little time for interaction during class. At school, teachers' observation of children's EF performance are largely based on students' behaviors in class, i.e., a relatively constrained setting that may not provide much opportunities for children to encounter emotion-elicited events that require emotional control. By contrast, parents should have more opportunities to interact with their children in less formal and restricted settings. Differential circumstances in functional settings in which the parents and teachers observed the children might explain why the correlation of the ERI reported by these two types of informants was nonsignificant.

In some Western countries (e.g., the United States), children in the elementary grades usually stay in the same classroom with a single teacher for all subjects, and this enables the teachers to spend a considerable amount of time with their students in class. In Hong Kong, however, students in a class are taught by at least 5-6 teachers throughout the day who rotate from one class to another, and teachers have little contact time with individual students, especially in a class of 30-40 people. We postulated that the limited interaction between teachers and students, along with the higher degree of regulation and constraint that characterized a typical classroom in Hong Kong, might provide another plausible explanation for the much lower interrater agreement between the BRIEF2 parent and teacher ratings in the Hong Kong sample versus those reported in other studies (Gioia et al., 2015).

Besides noticeable differences in the parent and teacher ratings on the BRIEF2, we also found dissimilar patterns of correlations between the two BRIEF2 forms and the children's performance on EF tasks. Ratings on the BRIEF-P correlated more with the working memory task scores, while the BRIEF-T correlated more with performance on tasks that measured shifting, attention, and inhibitory control. But more importantly, significant associations were not consistently revealed between performance on an EF task and the rating of the corresponding subscale on the BRIEF2 purportedly measuring the same EF construct.

Our findings corroborate those of prior studies that found only weak correlations between the BRIEF ratings and task-based performance of EF (Gerst et al., 2017; Toplak et al., 2009, 2013). One plausible explanation is that the assessment tasks and behavioral ratings are indeed measuring EF at two different cognitive functioning levels. As suggested by Toplak et al. (2013), ratings of EF based on behavioral observations are more indicative of one's abilities and motivation in the pursuit of one's goals, and thus are more related to the reflective level of cognitive functioning (Bratman et al., 1991). Whereas performance on an EF task may reflect more on the efficiency of EF processes at the algorithmic level without necessarily involving rational goal pursuit (Salthouse et al., 2003).

For instance, the Children's Color Trails Test is designed to assess cognitive flexibility (Llorente et al., 2003) by gauging the processing efficiency of the participant in shifting between sets of input. The Shifting subscale on the BRIEF2, by contrast, measures an individual's resistance to changes in daily routines, activities, and ways of tackling problems. Similarly, the Tower of London Test purportedly assesses an individual's ability to plan a course of action in a problem-solving task (i.e., to generate a sequence of moves to produce a configuration specified in the task). The abilities measured by this EF task might be contextually different from the behaviors exemplified in the Plan/Organize subscale of the BRIEF2, which are more relevant to the planning and completion of assignments and tasks that encompass more extensive goals.

Lastly, we examined how well the BRIEF2 ratings correlated with ADHD symptoms measured on the SNAP-IV. We considered the severity (i.e. the raw scores) as well as the number of symptoms to examine how well the BRIEF2 ratings correlated with the diagnostic symptoms of ADHD measured by the SNAP-IV. A strong correlation between the BRIEF2 and the SNAP-IV ratings would suggest high concurrent validity of the instrument in the identification of children with ADHD.

We found good convergent validity between the BRIEF2 and the SNAP-IV, as all the indexes of the BRIEF-P and the BRIEF-T were significantly correlated with the severity and the number of ADHD symptoms on the SNAP-IV when rated by the same informants. Much weaker associations were observed across different raters. Nonetheless, the CRI correlated significantly with IA and the BRI correlated with HI even across informants. The hierarchical regression analyses also indicated that both the BRIEF-P CRI and the BRIEF-T CRI contributed unique variance in the prediction of inattention over and above all other indexes. Similarly, the BRI on both the BRIEF-P and the BRIEF-T significantly predicted the HI symptoms on the SNAP-IV.

These results may have important implications for the interpretation of the BRIEF2 scores. More specifically, children rated with higher scores on the CRI subscales (i.e., doing worse in cognitive regulation) may likely display more significant attention problems, whereas those rated high on the BRI subscales may likely display stronger hyperactivity and impulsivity.

There are a couple of limitations to this study. First, except for working memory which was assessed on two subtests, we only administered one assessment task each to measure the EF construct of planning and problem solving, shifting, and attentional and inhibitory control. Although we were mindful that the reliability in the measurement of the construct might be discounted when based on a single indicator, this arrangement was primarily due to the practical constraints in logistics (i.e., not taking up too much of students' learning time at school) and also due to the consideration to avoid fatigue in the participants. Moreover, this study presented concurrent data based on evaluation at one time point. Future research may consider addressing this issue in a prospective study, for instance to collect data on children's ADHD symptoms at later time points to examine the predictive validity of the BRIEF2 in association with ADHD screening and identification.

Nonetheless, this study provided the first evidence for the factorial validity of the three-factor nine-scale model of the BRIEF2 Parent and Teacher Forms in an Asian sample of children with ADHD-related concerns. Results regarding the equivalence of the two forms and their correlations with performance on EF tasks highlight the importance of employing multiple methods based on multiple sources of information in multiple settings for neuropsychological assessments. The good convergent validity of the BRIEF2 with measures of ADHD symptoms (i.e., SNAP-IV) confirms its usefulness as an indicator of EF difficulties in children with ADHD concerns.

Importantly, the similarities in findings—in terms of factor structure, correlations with symptom rating scales, and limited correlations with performance-based measures—between the current study and previous research suggest that the BRIEF2 operates in a Hong Kong sample of children and adolescents much the same way as it does in Western populations. Our findings support the validity of the BRIEF2 in different cultural contexts and provide evidence for the cross-cultural stability of the EF construct. The poor interrater agreement between Hong Kong parents and teachers on the BRIEF2 in contrast to the more moderate agreement observed in clinical samples in the United States was perhaps the most salient difference between our results and those of prior studies. We postulate that disparities in the typical classroom setting in Hong Kong versus that in the United States might partially explain the results. Future research may further explore this difference to investigate whether other cultural factors are involved.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Erlbaum.

Anderson, P., Anderson, V., & Lajoie, G. (1996). The tower of London test: Validation and standardization for pediatric populations. *The Clinical Neuropsychologist, 10*(1), 54-65. https://doi.org/10.1080/13854049608406663

Antshel, K. M., Zhang-James, Y., Wagner, K. E., Ledesma, A., & Faraone, S. V. (2016). An update on the comorbidity of ADHD and ASD: a focus on clinical management. *Expert Review of Neurotherapeutics, 16(*3), 279-293. https://doi.org/10.1586/14737175.2016.1146591

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65–94. https://doi.org/10.1037/0033-2909.121.1.65

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved. Executive functions: What they are, how they work, and why they evolved.* Guilford Press.

Bausela-Herreras, E. (2018). BRIEF-P: analysis of executive functions according to informant / BRIEF-P: análisis de las funciones ejecutivas en función del informante. *Estudios de Psicologia*, *39*(2–3), 503–547. https://doi.org/10.1080/02109395.2018.1507096

Bayliss, D. M., & Roodenrys, S. (2000). Executive processing and attention deficit hyperactivity disorder: An application of the supervisory attentional system. *Developmental Neuropsychology*, *17*(2), 161-180. https://doi.org/10.1207/S15326942DN1702_02

Boivin, M. J., & Giordani, B. (2009). Neuropsychological assessment of African children:

evidence for a universal brain/behavior omnibus within a coconstructivist paradigm. *Progress in Brain Research*, *178*, 113-135. https://doi.org/10.1016/S0079-6123(09)17808-1

Bratman, M.E., Israel, D.J., & Pollack, M.E. (1991). Plans and resource-bounded practical reasoning. In J. Cummins & J. Pollock (Eds.), *Philosophy and AI: Essays at the interface (pp. 7–22)*. MIT Press.

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Cohen, M. J. (1997). *Children's Memory Scale.* The Psychological Corporation.

Conners, C. K. (2014). *Conners Continuous Performance Test 3rd Edition (Conners CPT 3).* Multi-Health Systems, Inc.

Conners, C. K. (2015). *Conners Kiddie Continuous Performance Test 2nd Edition (K-CPT 2).* Multi-Health Systems, Inc.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of London^{DX}: A standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology*, *13*(3), 285–301. https://doi.org/10.1093/arclin/13.3.285

Davidson, F., Cherry, K., & Corkum, P. (2016). Validating the Behavior Rating Inventory of Executive Functioning for children with ADHD and their typically developing peers. *Applied Neuropsychology: Child*, *5*(2), 127–137. https://doi.org/10.1080/21622965.2015.1021957

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Egeland, J., & Fallmyr, Ø. (2010). Confirmatory factor analysis of the behavior rating inventory of executive function (BRIEF): Support for a distinction between emotional and behavioral regulation. *Child Neuropsychology*, *16*(4), 326-337. https://doi.org/10.1080/09297041003601462

Elosúa, M. R., Olmo, S., & Contreras, M. J. (2017). Differences in executive functioning in children with Attention Deficit and Hyperactivity Disorder (ADHD). *Frontiers in Psychology, 8*(JUN), 1–11. https://doi.org/10.3389/fpsyg.2017.00976

Fournet, N., Roulin, J. L., Monnier, C., Atzeni, T., Cosnefroy, O., Le Gall, D., & Roy, A. (2015). Multigroup confirmatory factor analysis and structural invariance with age of the Behavior Rating Inventory of Executive Function (BRIEF)-French version. *Child Neuropsychology*, *21*(3), 379-398. https://doi.org/10.1080/09297049.2014.906569

Gau, S. S. F., Lin, C. H., Hu, F. C., Shang, C. Y., Swanson, J. M., Liu, Y. C., & Liu, S. K. (2008a). Psychometric properties of the Chinese version of the Swanson, Nolan, and Pelham, version IV scale-Teacher Form. *Journal of Pediatric Psychology, 34*(8), 850-861. https://doi.org/10.1093/jpepsy/jsn133

Gau, S. S. F., Shang, C. Y., Liu, S. K., Lin, C. H., Swanson, J. M., Liu, Y. C., & Tu, C. L. (2008b). Psychometric properties of the Chinese version of the Swanson, Nolan, and Pelham, version IV scale–parent form. *International Journal of Methods in Psychiatric Research, 17*(1), 35-44. https://doi.org/10.1002/mpr.237

Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(8), 1036-1045. https://doi.org/10.1097/00004583-199708000-00011

Gerst, E. H., Cirino, P. T., Fletcher, J. M., & Yoshida, H. (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. *Child Neuropsychology*, *23*(4), 381-407. https://doi.org/10.1080/09297049.2015.1120860

Gilotty, L., Kenworthy, L., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive skills and executive function in Autism Spectrum Disorders. *Child Neuropsychology*, *8*(4), 241-248. https://doi.org/10.1076/chin.8.4.241.13504

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

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *BRIEF-2, Behavior Rating Inventory of Executive Function (2nd Ed.).* Psychological Assessment Resources, Inc.

Halvorsen, M., Mathiassen, B., Amundsen, T., Ellingsen, J., Brøndbo, P. H., Sundby, J., Steinsvik, O. O., & Martinussen, M. (2019). Confirmatory factor analysis of the behavior rating inventory of executive function in a neuro-pediatric sample and its application to mental disorders. *Child Neuropsychology*, *25*(5), 599–616. https://doi.org/10.1080/09297049.2018.1508564

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huizinga, M., & Smidts, D. P. (2011). Age-related changes in executive function: A normative study with the dutch version of the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology*, *17*(1), 51–66. https://doi.org/10.1080/09297049.2010.509715

Jacobson, L. A., Kalb, L. G., & Mahone, E. M. (2019). When theory met data: factor structure of the BRIEF2 in a clinical sample. *The Clinical Neuropsychologist*, *34*(1), 243–258. https://doi.org/10.1080/13854046.2019.1571634

Jacobson, L. A., Pritchard, A. E., Koriakin, T. A., Jones, K. E., & Mahone, E. M. (2016). Initial Examination of the BRIEF2 in Clinically Referred Children With and Without ADHD Symptoms. *Journal of Attention Disorders*, 108705471666363. https://doi.org/10.1177/1087054716663632

Jiménez, A., & Lucas-Molina, B. (2019). Dimensional structure and measurement invariance of the BRIEF-2 across gender in a socially vulnerable sample of primary school-aged children. *Child Neuropsychology, 25*(5), 636–647. https://doi.org/10.1080/09297049.2018.1512962

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2019). semTools: Useful tools for structural equation modeling. R package version 0.5-2. https://CRAN.R-project.org/package=semTools

Lawrence, V., Houghton, S., Douglas, G., Durkin, K., Whiting, K., & Tannock, R. (2004). Executive function and ADHD: A comparison of children’s performance during neuropsychological testing and real-world activities. *Journal of Attention Disorders, 7*(3), 137–149. https://doi.org/10.1177/108705470400700302

Llorente, A. M., Voigt, R. G., Williams, J., Frailey, J. K., Satz, P., & D’Elia, L. F. (2009). Children's Color Trails Test 1 & 2: test–retest reliability and factorial validity. *The Clinical Neuropsychologist*, *23*(4), 645-660. https://doi.org/10.1080/13854040802427795

Llorente, A. M., Williams, J., Satz, P., & D’Elia, L. F. (2003). *Children’s Color Trails Test professional manual*. Psychological Assessment resources.

Lyons Usher, A. M., Leon, S. C., Stanford, L. D., Holmbeck, G. N., & Bryant, F. B. (2016). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Functioning (BRIEF) in children and adolescents with ADHD. *Child Neuropsychology, 22*(8), 907-918. https://doi.org/10.1080/09297049.2015.1060956

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., Singer, H. S., & Denckla, M. B. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology, 17*(7), 643-662. https://doi.org/10.1093/arclin/17.7.643

Mares, D., McLuckie, A., Schwartz, M., & Saini, M. (2007). Executive function impairments in children with attention-deficit hyperactivity disorder: Do they differ between school and home environments? *The Canadian Journal of Psychiatry, 52*(8), 527-534. https://doi.org/10.1177%2F070674370705200811

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function?. *Journal of the International Neuropsychological Society, 16*(3), 495-505. https://doi.org/10.1017/S1355617710000093

McCandless, S., & O'Laughlin, L. (2007). The clinical utility of the Behavior Rating Inventory of Executive Function (BRIEF) in the diagnosis of ADHD. *Journal of Attention Disorders, 10*(4), 381–389. https://doi.org/10.1177/1087054706292115

Peters, C., Algina, J., Smith, S. W., & Daunic, A. P. (2012). Factorial validity of the Behavior Rating Inventory of Executive Function (BRIEF)-Teacher form. *Child Neuropsychology, 18*(2), 168–181. https://doi.org/10.1080/09297049.2011.594427

Pino Muñoz, M., & Arán Filippetti, V. (2019). Confirmatory Factor Analysis of the BRIEF-2 Parent and Teacher Form: Relationship to Performance-Based Measures of Executive Functions and Academic Achievement. *Applied Neuropsychology: Child, 0*(0), 1–15. https://doi.org/10.1080/21622965.2019.1660984

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., & Todd, R. D. (2010). Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: implications for DSM-V and ICD-11. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(3), 217-228. https://doi.org/10.1016/j.jaac.2009.11.011

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 1.8.12.

Rezazadeh, S. M., Wilding, J., & Cornish, K. (2011). The relationship between measures of cognitive attention and behavioral ratings of attention in typically developing children. *Child Neuropsychology*, *17*(2), 197-208. https://doi.org/10.1080/09297049.2010.532203

Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02/

Rubia, K. (2011). "Cool" inferior frontostriatal dysfunction in attention-deficit/hyperactivity disorder versus "hot" ventromedial orbitofrontal-limbic dysfunction in conduct disorder: a review. *Biological Psychiatry, 69*(12), e69-e87. https://doi.org/10.1016/j.biopsych.2010.09.023

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.132.4.566

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 298*(1089), 199-209. https://doi.org/10.1098/rstb.1982.0082

Sjöwall, D., Roth, L., Lindqvist, S., & Thorell, L. B. (2013). Multiple deficits in ADHD: Executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *Journal of Child Psychology and Psychiatry, 54*(6), 619–627. https://doi.org/10.1111/jcpp.12006

Sonuga-Barke, E. J. S., Sergeant, J. A., Nigg, J., & Willcutt, E. (2008). Executive dysfunction and delay aversion in attention deficit hyperactivity disorder: Nosologic and diagnostic implications. *Child and Adolescent Psychiatric Clinics of North America, 17*(2), 367–384. https://doi.org/10.1016/j.chc.2007.11.008

Soreni, N., Crosbie, J., Ickowicz, A., & Schachar, R. (2009). Stop signal and Conners' Continuous Performance tasks: Test—retest reliability of two inhibition measures in ADHD children. *Journal of Attention Disorders*, *13*(2), 137-143. https://doi.org/10.1177%2F1087054708326110

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*(1), 53–72. https://doi.org/10.1080/09297040802070929

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability (WNV)*. Pearson.

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*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006

Williams, J., Rickert, V., Hogan, J., Zolten, A. J., Satz, P., D'Elia, L. F., Asarnow, R. F., Zaucha, K., & Light, R. (1995). Children's color trails. *Archives of Clinical Neuropsychology*, *10*(3), 211-223. https://doi.org/10.1016/0887-6177(94)00041-N

Zelazo, P. D., and Müller, U. (2002). "Executive function in typical and atypical development," in *Blackwell Handbook of Childhood Cognitive Development*, ed U. Goswami (Malden, MA: Blackwell Publishers Ltd.), 445–469.

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354-360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Qu, L., & Muller, U. (2004). Hot and cool aspects of executive function: Relations in early development. In W. Schneider, R. Schumann, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive Junctioning, working memory, verbal ability, and theory of mind (pp. 71–93)*. Lawrence Erlbaum Associates Publishers.