Barkley Deficits in Executive Functioning Scale (BDEFS): Validation in a large multisite college sample

Jaclyn M. Kamradt[1], **Molly A. Nikolas**[1], **G. Leonard Burns**[2], **Annie A Garner**[3], **Matthew A. Jarrett**[4], **Aaron M. Luebbe**[5], **Stephen P. Becker**[6,7]

[1]Department of Psychological and Brain Sciences, University of Iowa

[2]Department of Psychology, Washington State University

[3]Department of Psychology, St. Louis University

[4]Department of Psychology, University of Alabama

[5]Department of Psychology, Miami University

[6]Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center

[7]Department of Pediatrics, University of Cincinnati College of Medicine

Abstract

Despite the importance of daily life executive functioning (EF) for college students' success, few measures exist that have been validated in college students specifically. This study examined the factor structure of the Barkley Deficits in Executive Functioning Scale (BDEFS) in college students. Participants were 1,311 students (ages 18–28 years, 65% female) from five universities in the United States. Additionally, the study examined invariance across sex, age, and attention-deficit/hyperactivity disorder (ADHD) symptoms. Exploratory structural equation modeling (ESEM) provided strong support for the BDEFS five factor structure though some items had high cross-loadings on multiple factors. Findings generally supported invariance across sex and age; however, loadings, thresholds, and factor means differed based on ADHD symptoms. Stronger support for invariance across sex emerged for a reduced item version that eliminated cross-loading items. Overall, findings provide support for the validity and utility of the BDEFS in college students.

Keywords

ADHD; emerging adults; executive functions; psychometrics; sex differences; university

Correspondence concerning this article should be addressed to Jaclyn M. Kamradt, Department of Psychological and Brain Sciences, W311 Seashore Hall, University of Iowa, Iowa City, IA 52242. jaclyn-kamradt@uiowa.edu, Phone: 319-335-2406, Fax: 319-335-0191.

Author Manuscript

College success is robustly related to individual differences in executive functioning (EF), which include a set of component cognitive abilities involved in strategic planning, cognitive flexibility, self-regulation, and goal-directed behavior (Lezak, 1995; Rabin, Fogel, & Nutter-Upham, 2011; Weyandt, 2005). Demands for executive skills are high in college, as students are expected to work more independently than in previous points in development, requiring increased skills in organization, time-management, and planning. In fact, recent work has found that teaching these skills to college students may decrease academic impairment (LaCount, Hartung, Shelton, & Stevens, 2018). Although sizeable research has examined EF in relation to psychopathology, including ADHD, autism spectrum disorders, and mood disorders (O'Hearn, Asato, Ordaz, & Luna, 2008; Snyder, 2013; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005), it is important to note that EF deficits can be present and problematic regardless of clinical status (Biederman et al., 2006). For example, Biederman and colleagues (2006) found that EF deficits in adults were associated with lower academic achievement, independent of ADHD status. Moreover, EF is relatively stable over time and has a broad impact across the life span, as early life self-control is associated with adult outcomes including education, health, income, and crime (Friedman et al., 2016; Friedman et al., 2008; Mischel et al., 2011; Moffitt et al., 2011; Velez-Pastrana et al., 2016). Thus, assessing EF deficits in a population of young adults (e.g., college students), may identify individuals at risk for associated impairments in important life domains and highlight intervention targets (e.g., poor organization, lack of self-restraint) to help prevent or remediate such outcomes.

Importantly, EF is especially relevant for college students as they are faced with a multitude of academic and social demands that require key aspects of EF (e.g., goal setting) that are critical for their success (Munro, Weyandt, Marraccini, & Oster, 2017). Unfortunately, many college students have difficulties with setting goals (e.g., setting too few or inappropriate goals) and have ineffective self-regulatory skills that are necessary to achieve those goals (Schutz, White, & Lanehart, 2000). Additionally, well-established self-regulatory skills seem crucial for college students, as they regularly need to engage in self-generated thoughts, feelings, and behaviors that are intended to meet goals and achieve future success. Importantly, failure to set clear academic goals is associated with course dropout and even college dropout rates (Morisano, Hirsh, Peterson, Pihl, & Shore, 2010).

Relatedly, procrastination, or the intentional delay of due tasks, is another common problem among college students that may be in part due to deficits in EF (Rabin et al., 2011). For example, Rabin and colleagues (2011) demonstrated that multiple deficits in self-reported mental processes related to EF, including initiation, planning/organizing, inhibition, self-monitoring, working memory, task monitoring, and organization of materials, were significant predictors of academic procrastination among college students. Notably, up to 60% of undergraduate students report procrastination in completing school tasks that, in turn, negatively impacts their grades (Kachgal, Hansen, & Nutter, 2001; Onwuegbuzie, 2004). Further, chronic procrastination can result in poor performance on educational tasks, decreased learning, increased health risks, and relationship impairments (e.g., strain on relationships) (Burns, Dittman, Nguyen, & Mitchelson, 2000; Moon & Illingworth, 2005; Tice & Baumeister, 1997). Taken together, individual differences in EF among college students appear to be important for predicting both academic performance and retention

(Petersen, Lavelle, & Guarino, 2006; Schutz et al., 2000). Consequently, assessment of these differences may be useful for understanding college learning and identifying modifiable factors that can improve important academic outcomes (e.g., retention rates). Therefore, researchers, clinicians, and career advisors need tools that are valid and reliable indicators of the EF construct in college populations. However, to date, psychometric studies on EF rating scales among college students have been limited.

Measurement of Daily Life EF in College Students: Barkley Deficits in Executive Functioning Scale (BDEFS)

According to Barkley, EF involves two higher-order factors comprised of multiple lower-order domain skills: inhibition (e.g., ability to inhibit motor, verbal, cognitive, and emotional responses) and metacognition (e.g., nonverbal working memory, verbal working memory, planning, problem-solving, emotional self-regulation) (Barkley, 2010). These domains have traditionally been assessed via two methods: EF tasks and EF rating scales. However, the mixed findings and low convergence of EF tasks and EF ratings has raised questions about the optimal way to assess EF in young adults. For example, Barkley (2010) noted there may be added ecological validity of EF rating scales versus tasks. Other work has demonstrated little overlap between EF tasks and ratings, particularly self-report ratings of EF (Barkley & Fischer, 2011; Bogod, Mateer, & MacDonald, 2003; Toplak, West, & Stanovich, 2013). Further research has indicated that EF ratings may offer additional predictive value of functional impairment (over and above that of EF tasks), whereas EF tasks may be closely associated with severity of psychopathology symptoms (Kamradt, Ullsperger, & Nikolas, 2014).

Although there are several rating scales proposed to assess EF in adulthood, such as the Behavior Rating Inventory of Executive Function-Adult Version (Roth, Isquith, & Gioia, 2005), the Dysexecutive Questionnaire (Burgess, Alderman, Evans, Emslie, & Wilson, 1998) and the Executive Function Index (Spinella, 2005), the utility of these scales has been limited by low reliability of scores and inconsistent factor structures (Gerstorf, Siedlecki, Tucker-Drob, & Salthouse, 2008; Janssen, De Mey, & Egger, 2009; Velez-Pastrana et al., 2016). In response to those limitations, the Barkley Deficits in Executive Functioning Scale (BDEFS; Barkley 2011), based on a more representative U.S. sample than prior measures, has emerged as a reliable measure of EF (internal consistency alpha coefficients in manual reported at >0.91), with some evidence of ecological validity (Barkley, 2011). In addition to its potential relevance to psychopathology (Barkley, 2011; Barkley, 2012; Jarrett, 2016), measures of EF, such as the BDEFS (Barkley, 2011), may have utility in career planning among college students (Prevatt & Yelland, 2015). Recently, an additional scale, the Comprehensive Executive Function Inventory-Adult, has been published (Naglieri & Goldstein, 2017); however additional work evaluating its construct validity has not yet been published. Overall, few studies have examined the validity of score interpretations based on EF measures among college students specifically, which is needed to determine the utility of these measures.

Given the aforementioned advantages of using the BDEFS specifically in assessment of EF of college students, it is important to understand its original structure in the populations for which it was intended. The BDEFS measure includes 89 items that assess multiple components of EF, including five separable problem domains described as Self-Management to Time, Self-Organization/Problem Solving, Self-Restraint, Self-Motivation, and Self-Regulation of Emotion (Barkley, 2011). The basis of these five factors stems from the principal components factor analysis that was applied to 91 EF items on the Deficits in Executive Functioning Scale from Barkley's original item pool based on EF theory of ADHD (Barkley, 1997; Barkley & Murphy, 2011). Consequently, five factors emerged and were retained based on their Eigenvalues and other model fit criteria (see Barkley, 2011). Of the 91 items, three items were dropped from the scale because they did not meet the factor loading threshold (.40) on any of the five final factors in the analysis, resulting in the final 89-item, five factor scale (Barkley, 2011). While the scale was originally developed and validated to evaluate EF deficits in a nationally representative sample of adults aged 18–81 years (Barkley, 2011), less work has focused on the utility of this measure in a college specific population. Examining daily life EF in the general population is important since both ADHD and EF are dimensional in nature (Carragher et al., 2014; Larsson, Anckarsater, Rastam, Chang, & Lichtenstein, 2012; Marcus, Norris, & Coccaro, 2012). The validity of this measure in a college student population remains an open question given the little work in this area.

Despite its potential clinical utility, the structure and validity of the BDEFS has yet to be examined in an English-speaking college population. One study examining an intervention of ADHD coaching in college students concluded that students commonly identified problem areas that aligned consistently with Barkley's five domains of EF (Prevatt & Yelland, 2015). Interestingly, recent factor analytic work examining the validity of a Spanish version of the BDEFS in a community sample (i.e., college students and other adults) supported the five-factor model of the BDEFS, demonstrating the Spanish version in college students and adults seems to assess similar dimensions of EF as the original English version (Velez-Pastrana et al., 2016). However, eleven items (of the 89 retained in the BDEFS) did not load significantly nor similarly on both versions (English- and Spanish-speaking versions) and were removed in the final confirmatory models (Velez-Pastrana et al., 2016). Overall, this study provides important preliminary insight into the factor structure of the BDEFS in a sample that *includes* college students; however, factor analytic work is still needed to validate this measure in an English-speaking, *all* college population.

Present Study

Research evaluating the validity of the BDEFS scores among young adults is underway. One prior study evaluated its criterion validity by examining how much the BDEFS scales add to the prediction of ADHD symptoms compared to traditional executive function tests (Dehili, Prevatt, & Coffman, 2017). However, no study to date has investigated the construct validity of the BDEFS measure in a large, English-speaking college population, leaving a critical gap in knowledge regarding how to best assess EF in college students. Therefore, the purpose of the present research is to determine the validity of the Barkley Deficits in Executive Functioning Scale (BDEFS) in a college population. This will be done by (1) comparing

measurement models, including exploratory structural equation modeling (ESEM) and confirmatory factor analysis (CFA), to identify the optimal, parsimonious factor structure of the BDEFS in a large multi-site college sample, and (2) examining the invariance of the BDEFS measurement model across sex, age, and levels of ADHD symptoms.

Evaluation of invariance across sex and age are particularly important, given that there is evidence of potential sex differences in EF beginning in childhood (Wodka et al., 2008), which may also impact important changes in EF abilities across development during early adulthood (Huizinga, Dolan, & van der Molen, 2006). Further, we also elected to examine the invariance of the BDEFS across levels of ADHD symptoms, given that EF deficits have often been associated with ADHD (Barkley, 1997, 2010; Jarrett, 2016). Overall, we predict that the five-factor structure of the BDEFS will emerge within a college population, similar to prior work with nationally representative adult samples (Barkley, 2011; Barkley, 2012). We also predict that validity analyses will reveal evidence of invariance across sex and age, but that ADHD symptoms will likely impact the BDEFS factor means, given prior work.

Methods

Participants

Participants included 1,311 undergraduate students enrolled in five universities in the United States[1]. The majority of the universities are public universities (four out of the five) and they are located in the Midwestern, Southeast, and Northwest regions of the United States. Participants were 18 to 28 years (M = 19.30 years, SD = 1.44) and approximately two thirds were female (64.9%, n=851). This sample's age and sex approximately matches that of the US college student population (US Department of Education, 2018). Participants' self-reported race/ethnicity and year in school is reported in Table 1. The majority of particiapnts identified as White (79.8%). Most participants were in their 1st year of college (64.1%). Although a formal psychiatric assessment/diagnosis was not conducted in the context of this multisite study, participants reported if they had ever received a professional mental health diagnosis, with three-quarters of participants (n=979) reported never receiving a mental health diagnosis (see Table 1).

Procedures

This study was approved by the local Institutional Review Board (IRB) at each university, with the individual study protocols specifying that data would be merged across sites for analysis and dissemination. Procedures varied slightly based on normative practices at each institution[2]. At four of the sites, this study was an anonymous online survey. Specifically, after signing up for the study via an online system, participants were directed to an online survey in Qualtrics where they first read an information sheet describing the study and

---

[1]Participants were from a larger study (n=3,172) that unfortunately included n=1,861 participants with missing data on the final three BDEFS items (items 87, 88, and 89) due to an administration error. Thus, these data were not considered to be missing at random and these participants were excluded from these analyses. Importantly, to ensure that the remaining sample that resulted in n=1,311 participants was similar to those who were excluded, full descriptive information on the full and resulting reduced sample is provided (see Supplemental Table 6). Although eliminating those systematic missing data results in a smaller sample, the sample was still large enough to compare models with confidence.
[2]Participants did not differ in meaningful ways based on site. Site was generally unassociated with EF factor scores (all rs<|.06|), with no significant correlations to note.

providing contact information of the local investigator, IRB, and student counseling center. If the participant chose to continue, they were then directed to the survey, and after completing the survey, automatically received course credit for their participation. At the fifth university, participants were given an individual time-slot for coming to the investigator's laboratory, and after providing informed consent in-person, completed the same Qualtrics survey as participants at the other four universities on their own time. Course credit was granted similarly across participants.

Measures

Sociodemographic characteristics were collected via a self-report questionnaire. Data included age, sex, race/ethnicity, year in school, parental educational attainment, parental income, and psychiatric diagnoses.

Executive Functioning.—The Barkley Deficits in Executive Functioning Scale (BDEFS; Barkley, 2011) was used to assess daily life EF. The BDEFS includes 89 items that assess multiple EF domains, including Self-Management to Time (e.g., "have trouble doing what I tell myself to do"; Cronbach's α = .96), Self-Organization/Problem Solving (e.g., "have trouble doing things in their proper order or sequence"; α =.96), Self-Restraint (e.g., "likely to do things without considering the consequences for doing them"; α =.93), Self-Motivation (e.g., "I do not have the willpower or determination that others seem to have"; α =.93), and Self-Regulation of Emotion (e.g., "overreact emotionally"; α =.93). Internal consistences were computed based on item-level data in the current sample. Participants rated each item on a four-point scale (1 *never or rarely*, 4 *very often*) in reference to the past six months, with higher scores indicating greater EF deficits. The BDEFS has demonstrated satisfactory 2- to 3-week test–retest reliability and satisfactory validity for measuring EF in a national adult sample (Barkley, 2011). For correlations, means, and standard deviations of BDEFS factors, see supplementary table 7.

ADHD symptoms.—The Barkley Adult ADHD Rating Scale-IV (BAARS-IV; Barkley, 2011a) was used to assess ADHD symptoms. The BAARS-IV includes 18 items that are consistent with the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) symptoms of ADHD that have been updated in their wording to also reflect modifications made in *DSM-5* (American Psychiatric Association, 2013). Using a four-point scale (in this study, 0=*not at all*, 3=*very often*), participants respond to each item with reference to how often each statement best describes their behavior over the past six months. The ADHD-IN (e.g., "difficulty sustaining my attention in tasks or fun activities") and ADHD-HI (e.g., "fidget with hands and feet or squirm in seat") subscales of the BAARS-IV have demonstrated satisfactory internal consistency and test–retest reliability over a 2- to 3-week time period (Barkley, 2011). In the present study, Cronbach's αs were .89 and .83 for the ADHD-IN and ADHD-HI dimensions, respectively.

Analytic strategy

This study aimed to (1) identify the best fitting measurement model of BDEFS using both Exploratory Structural Equation Modeling (ESEM) and Confirmatory Factor Analysis

(CFA), and (2) examine measurement invariance of the BDEFS in a college population. Given that the current data are derived from questionnaires with a 4-point Likert scale response system, and so are ordinal (not continuous), all models in the current analysis were run using robust weighted least squares mean and variance estimator (WLSMV) in *Mplus* (Muthen & Muthen, 2011). Missingness did not exceed 0.70% for any item on the BDEFS in this sample. Exploratory factor analyses were not employed here, given past work has consistently demonstrated a five-factor solution for this measure. The analysis occurred in the following stages.

Prior to running primary analyses, data were screened for invalid responses (see Becker et al. 2018 for a thorough description of these procedures). To improve the quality of participant responses, an instructional manipulation check (IMC; Oppenheimer et al. 2009) was used to measure whether participants read the instructions carefully. "Trap" questions were also used to detect individuals who were quickly responding to survey questions without sufficient attention to item content. In addition, one question was included at the end of the full survey that asked participants the following: "How much effort did you put into this study from 0 to 10 (0 = not much effort at all, 5 = moderate effort, 10 = my best effort)?" To ensure the validity of responses, a threshold was set of 50% accuracy or higher for the "trap questions" and a self-reported effort rating of 5 or higher. In order for data to be included in this sample, participants had to answer 50% of trap questions correctly and have an overall effort of 5 or higher (Becker et al., 2018).

Stage 1 was conducted to evaluate the factor structure of the BDEFS. First, we utilized ESEM and CFA to examine the fit of the five-factor BDEFS measurement model. Given that prior work has identified five factors (Barkley, 2011; Velez-Pastrana et al., 2016), we elected to begin with ESEM (rather than EFA) and CFA to assess the fit of a five-factor model as well as to identify potential problematic items with high cross-loadings. In ESEM, the factor structure is specified; however, items are allowed to cross-load on other factors. Factors are also expected to be correlated, so oblique rotations were used. CFA models were specified in accordance with the original measure publications such that each item only loaded on one factor (simple structure). Recent research has suggested that using a combination of ESEM and CFA is the most fruitful approach to examining inventory structure at this time, as there is not enough evidence that ESEM is necessarily more advantageous than or should be used in place of CFA (Booth & Hughes, 2014). We generated two random samples from our full sample (n=655 and n=656) and used one to examine ESEM and one to examine CFA. Statistical comparison of demographic characteristics of both samples demonstrated that random assignment was adequate. Model fit was assessed according to cut-off values for the indices of absolute and relative model fit, which include values of >.95 for the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) (Hu & Bentler, 1998, 1999), and values of <.06 for the Root Mean Square Error of Approximation (RMSEA) (Browne & Cudeck, 1993). Next, we utilized results of the ESEM model to identify potential problematic items (i.e., items with high cross-loadings on multiple factors) and then tested the fit of various reduced-item model sets using CFA in the full sample. Because these models were non-nested, we descriptively compared the fit of each model based on the CFI, TLI, and RMSEA.

Stage 2 analyses utilized multiple group CFA in the full sample to evaluate measurement invariance across sex. First, we confirmed adequate fit of the five-factor model separately in males and females. We then proceeded to test invariance by fitting two models. First, we fitted a configural (or equal forms) model, which freely estimated the factor loadings and thresholds of the five-factor BDEFS model across males and females (i.e., the loadings and thresholds are allowed to freely vary across sex). Next, we fit a full threshold invariance model (or full scalar invariance model), which constrained both factor loadings and thresholds to be equivalent across sex. We completed these analyses in line with prior writing on invariance with ordinal and categorical items, which has suggested constraining both loadings and item thresholds in a single step rather than testing loading and threshold equivalence separately (Brown, 2015). We also examined modification indices (corrected for multiple tests) to probe potential sources of invariance across item loadings, thresholds, and latent factor means.

Given multiple ways of comparing models in invariance analyses, we utilized two approaches: chi-square difference tests (given use of WLSMV) and change in CFI (Little, 2013). First, we utilized the DIFFTEST option in Mplus to calculate the chi-square difference test, given that we utilized a WLSMV estimator in these models. While the chi-square difference is a widely-used standard for global evaluation of invariance, our large sample size and item set were substantially powered to detect small (and potentially less meaningful) differences across sex. Thus, we also employed the change in CFI, given prior work supporting its utility in evaluating the invariance of loadings and intercepts/thresholds as well as loadings and thresholds (Little, 2013). More recent work has suggested some potential issues when solely using the change in CFI in evaluating invariance in large samples; thus, we report both metrics here. All models used theta parameterization in order to generate fit diagnostic information (i.e., modification indices) and determine potential contributions to noninvariance (Brown, 2015). Using this information, tests of partial invariance were subsequently conducted. We repeated these analyses in a reduced item set; this removed items deemed "problematic" due to high cross-loadings.

Stage 3 analyses also evaluated measurement invariance and involved conducting a series of multiple indicator multiple cause (MIMIC) models. These models can be useful for evaluating invariance across relevant continuous variables (age, ADHD symptoms), which may preclude evaluation using the traditional multiple group CFA (MaCallum, Zhang, & Preacher, 2002). However, these models assume invariance across factor loadings. Therefore, we first utilized a multiple-group CFA approach to test for invariance in loadings among those low and high (determined by a median split) on a given covariate (age, inattention symptoms, hyperactivity-impulsivity symptoms). To expand on issues of invariance across sex, we also conducted MIMIC models for sex. To ensure adequate fit (form invariance), models were tested separately for males and females. We then proceeded to test measurement invariance. For these models, each of the latent factors and all 89 items were regressed on the relevant covariate (sex, age, inattention, hyperactivity-impulsivity) to determine its impact on the factor means. Per Brown (2015), models were first conducted with restricting the direct associations between each of the items and the covariate to zero and examining modification indices to determine items with differential functioning.

Results

Exploratory Structural Equation Model and Confirmatory Factor Analysis

Given the focus on the construct validity of the BDEFS and invariance across three key factors (age, sex, and ADHD symptoms), we first examined how the five scales of this measure varied across sex, age, and ADHD symptoms. Several differences based on sex emerged. Specifically, compared to males, females reported significantly more problems on the organization/problem-solving scale, $t(1308)=-2.11$, $p=0.04$, and on the self-regulation of emotions scale, $t(1309)=-4.34$, $p=0.00$; however, males reported significantly more problems on the self-restraint scale, $t(1307)=2.61$, $p=0.01$, and on the self-motivation scale, $t(1307)=2.78$, $p=0.01$. Age was significantly correlated with self-restraint and self-regulation of emotions, such that younger participants reported increased problems with restraint ($r=-0.06$, $p<.05$) and emotion regulation ($r=-0.07$, $p<.05$). Moreover, inattention symptoms were signicantly correlated with all five scales (rs= 0.43–0.61, all $p<.01$). Hyperactivity-impulsivity symptoms were also significantly correlated with all five scales (rs= 0.27–0.43, all $p<.01$). Additionally, all five scales were significantly correlated with each other (rs= 0.48–0.72, $p<.01$).

We next examined the fit indices of the five-factor ESEM and CFA models that were conducted in the two random samples (see Table 2). Although the CFA had acceptable fit, ESEM emerged as the optimal measurement model in these initial comparisons based on fit indices favoring the ESEM. The better relative fit of the ESEM model is not surprising, given that ESEM allows for multiple cross-loadings whereas CFA does not.

Overall, from the ESEM, we identified items as being potentially problematic if they had high cross-loadings on multiple factors (i.e., loaded on their original factor as well as had a high cross-loading on a second factor). We utilized a range of cross-loading thresholds (.25, .30, and .35) to identify potentially problematic items. Examining several thresholds allowed for broad exploration of items that may be problematic and for comparison of such items to another validation study that examined a Spanish version of the BDEFS (Velez-Pastrana et al., 2016). The following 20 items met criteria for high cross-loadings based on a threshold of .25 (e.g., item had a loading greater than >.25 on the factor to which it is assigned *and* another factor): items 5, 9, 10, 18, 23, 24, 26, 36, 39, 41, 43, 44, 46, 55, 56, 60, 63, 64, 69 and 77. Nine items had cross-loadings >.30 (items 5, 9, 10, 24, 36, 41, 43, 44, and 77), whereas only two items had cross-loadings >.35 (5 and 43).

To determine the best fitting model, using the full sample of participants (n=1,311) we compared the fit indices from the CFA model that included all of the BDEFS items to models that removed the aforementioned items at each of the indicated thresholds. Based on fit indices (CFI, TLI, and RMSEA), removing items with higher cross-loadings did not substantially improve model fit, and model fit was still adequate at each threshold. Given the comparable fit across models as well as prior work indicating that maintaining cross-loadings below .30 provides the "cleanest" factor structure (Costello & Osborne, 2005), we proceeded with invariance analyses using both the full measure as well as with removing items with cross-loadings greater than .30 (e.g., items 5, 9, 10, 24, 36, 41, 43, 44, 77 removed). Scale reliability composites for the latent factors were computed as omegas using

the factor loadings and error variance estimates for each factor (using the Maximum Likelihood with Robust Standard Errors estimator) and were adequate for each factor in the full item set (Factor 1, Self-Management to Time=.87; Factor 2, Organization/Problem-Solving=.90; Factor 3, Self-Restraint=.82; Factor 4, Self-Motivation=.91; Factor 5, Self-Regulation of Emotions=.88; Total BDEFS Score=.89) and in the reduced item set (Factor 1, Self-Management to Time=.81; Factor 2, Organization/Problem-Solving=.84; Factor 3, Self-Restraint=.79; Factor 4, Self-Motivation=.87; Factor 5, Self-Regulation of Emotions=.82; Total BDEFS Score=.86).

Invariance Across Sex

Full set of items.—The five-factor model of the BDEFS evidenced acceptable model fit in males and females separately, as did the configural model (or equal forms model), which allowed factor loadings and thresholds to be freely estimated across sex (see Tables 2 & 3). Next, scalar (or full threshold) invariance was evaluated by constraining the item factor loadings and the thresholds for each item to be equivalent across sex. The chi-square difference test revealed a significant reduction in fit (i.e., restrictions added to create the nested model result in significantly worse fit); however, the change in CFI reflected *improved* model fit of the constrained model (see Table 3). Examination of modification indices revealed potential sex differences in 12 factor loadings, in 35 thresholds across 21 different items, as well as significant sex differences in the latent factor means. A partial invariance model was then examined, which involved freeing these parameters. The chi-square difference test revealed a non-significant change relative to the equal forms model as well as improved model fit based on the change in CFI, suggesting evidence for partial invariance (see Tables S1–S5 for full list of item factor loadings overall and by sex).

Reduced set of items.—We next conducted invariance analyses across a smaller subset of items (e.g., removing those with cross-loadings >.30). Findings indicated a non-signfiicant change in chi-square in the full invariance model relative to the equal forms model and an improvement in CFI, providing evidence for invariance across sex in this reduced subset of items.

Multiple Indicators Multiple Causes Confirmatory Factor Analysis (MIMIC)

The effects of covariates (sex, age, inattention, hyperactivity-impulsivity) on factor means and scale items were examined respectively using MIMIC models. These models were conducted by first examining invariance in factor loadings within a multiple-group CFA framework (as MIMIC models assume invariance in loadings) followed by models that constrained the direct effects of each covariate on each item at zero and examining modification indices (Brown, 2015). Results from these models included all items from the BDEFS and are shown in Table 4.

Sex.—In addition to the multiple group CFA described above, MIMIC models for sex were also conducted. Four item thresholds varied by sex (see Table 4). Sex was negatively associated with self-restraint and self-motivation, such that males had more problems in these areas than females. In constrast, females had significantly higher means on self-regulation compared to males.

Age.—Multiple-group CFA models revealed invariance in item loadings across age groups (divided at age 19 and age 20). It is important to note that the five-factor model fit well in both younger (RMSEA=0.043, CFI=.933) and older participants (RMSEA=0.041, CFI=.939). In a multiple group CFA format, constraining factor loadings to be equivalent across older and younger participants did not result in a significant detriment in model fit (DIFFTEST χ^2=78.12, df=84, p=.66, CFI=.001). Age was modeled continuously in MIMC models. MIMIC models constraining all direct effects between the items and age revealed no significant direct effects of age on any of the items as all Modification Indices <15.00, (p-value corrected for multiple testing). Age was significantly inversely associated with the restraint factor (β=−.060, p=.048) and the self-regulation factor (β=−.062, p=.032), indicating that as age increased, scores on problems with restraint and problems with self-regulation decreased slightly (see Table 4).

Inattention.—The five factor model fit well in both those with low (RMSEA=0.038, CFI=.935) and high levels of inattention symptoms (RMSEA=0.040, CFI=.939). In a multiple group CFA format, constraining factor loadings to be equivalent across those with low and high levels of inattention symptoms resulted in a significant detriment in model fit (DIFFTEST χ^2=206.54, df=84, p<.001, CFI=.033), suggesting differences in item loadings across levels of inattention. MIMIC models revealed significant positive associations between all five factors and inattention. Additionally, inattention was also significantly associated with 6 items (after correction for multiple tests, MI>15.00), suggesting that those items may perform differently depending on the individual's level of inattention symptoms.

Hyperactivity-Impulsivity.—The five factor model fit well in both those with low (RMSEA=0.036, CFI=.943) and high levels of hyperactivity-impulsivity symptoms (RMSEA=0.041, CFI=.938). In a multiple group CFA format, constraining factor loadings to be equivalent across those with low and high levels of hyperactive-impulsive symptoms resulted in a significant detriment in model fit (DIFFTEST χ^2=187.44, df=84, p<.001, CFI=.038), suggesting differences in loadings across levels of hyperactivity-impulsivity. MIMIC models revealed significant positive associations between all five factors and hyperactivity-impulsivity. Additionally, 4 items also showed direct associations with hyperactivity-impulsivity, suggesting differential performance of those items.

Discussion

Executive function skills are crucial to college student success; however, research to date has left a gap in knowledge regarding the best way to assess EF in a college population. Identification of reliable and valid measures for assessing deficits in EF among college students could provide an important advance for determining those at risk for problematic outcomes, including chronic procrastination, nonmedical stimulant usage, poor grades, low educational attainment, and occupational difficulties. We set out to evaluate the validity of the Barkley Deficits in Executive Functioning Scale (BDEFS) as one such possible measure.

Initial models compared the fit of ESEM and CFA models within two random halves of the sample. The model fit using ESEM allowed for cross-loading of items and fit the data

somewhat better than the CFA model, suggesting some issues with item cross-loadings (9 items out of 89). Overall, results supported a five-factor structure but suggest the variation in some item responses may be influenced by multiple EF domains. This is not surprising, given that component processes of EF are often interrelated (Pennington & Ozonoff, 1996). Examination of items with high cross-loadings was also somewhat consistent with prior work by Velez-Pastrana and colleagues (2016), who examined a Spanish version of the BDEFS and identified 11 problematic items for removal (items 41, 43, 44, 46, 51, 56, 60, 61, 62, 64, and 69). Specifically, three problem items from their Spanish-language BDEFS study overlapped with problems items identified in the present study (at the >.30 threshold). These overlapping items included: item 4 (*Cannot focus my attention on tasks or work as well as others*), item 43 (*Can't seem to sustain my concentration on reading, paperwork, lectures, or work*), and item 44 (*Find it hard to focus on what is important from what is not important when I do things*). Additionally, when examining the more liberal threshold of >.25, we found striking consistency in problem items with the Spanish-language BDEFS study, with 8 out of the 11 items they removed found to be overlapping in our study (items 41, 43, 44, 46, 56, 60, 64, and 69). It should also be noted that although we retained items 61 and 62 in our examination of problem items at the >.25 threshold, these two items may still be problematic. Our ESEM results demonstrated that these two items have loadings less than .25 on Factor 3 (where they are assumed to load based on initial principal components analyses conducted in the development of the measure, see Barkley, 2011), but high (>.25) loadings on Factor 5. Notably, these items were removed in the Spanish BDEFS validity study. Overall, the identification of problem items at this level bare remarkable consistency (10/11 items overlapping) with the Spanish-language BDEFS study. Moreover, the present study is strengthened by applying more stringent methods of identifying items for removal, yet our findings still overlapped with previous work (Velez-Pastrana et al., 2016).

Despite potential problematic items due to high cross-loadings, the findings of the present study replicated the five-factor structure of the BDEFS in a non-clinical sample. Notably, we did not observe large correlated residuals or other indicators of major localized ill fit, providing further support for a five-factor solution. Removal of items with high cross-loadings did not result in significant improvements or detriments in fit and a five-factor framework was supported in all models. However, given potential item overlap as well as high cross-loadings, future work may benefit by utilizing parallel analysis or other simulation-based approaches to confirm the number of factors in future samples (Garrido, Abad, & Ponsoda, 2013; Timmerman & Lorenzo-Seva, 2011). Additionally, given that each factor is comprised of a relatively large number of items and similar items (e.g., items 13, 14, and 15 on the Self-Management to Time factor all assess motivation for starting, continuing, and completing work or activities), item parceling may be beneficial in future work utilizing structural models (Cole, Perkins, & Zelkowitz, 2016). Item parceling may also be indicated given the large number of correlated item residuals; however, this practice in CFA, particularly when evaluating invariance, remains controversial (Little, Rhemtulla, Gibson, & Schoemann, 2013; Marsh, Ludtke, Nagengast, Morin, & Von Davier, 2013).

Invariance testing generally supported invariance across sex. Importantly, given there is no "gold standard" for interpreting invariance analyses, we presented two approaches in evaluating the invariance across sex. When examining the change in CFI (i.e., Little, 2013),

we found evidence of invariance on loadings and thresholds across sex. However, in taking a more conservative approach by examining the chi-square differences, findings were more ambiguous, although there was some evidence supporting partial metric invariance (e.g., some equivalence in item factor loadings across sex). It is likely that our large sample size and the large item set contributed to detection of significant differences in loadings and thresholds that may not be meaningfully different (see Supplementary materials). Further, MIMIC models revealed significant sex effects on factor means as well as in thresholds for four items. Thus, it appears that at least some items from the BDEFS are performing differently for males and females (see Tables S1–S5). Future work may also benefit from examining items/factors that may be more relevant for predicting outcomes (e.g., procrastination, academic problems) differentially by sex.

We also evaluated invariance across age and ADHD symptoms using MIMIC models. In general, there was evidence supporting invariance across age, although it is important to point out the range restriction in our college student sample. Loadings and thresholds were largely equivalent across age. Further, age only appeared to impact the mean of the self-regulation factor. However, MIMIC models revealed that ADHD symptoms influenced some item thresholds as well as factor means. These results are not entirely surprising given that many EF items (e.g., "Easily distracted by irrelevant events or thoughts when I must concentrate on something") seem to have content overlap with ADHD symptoms items. Thus, our findings support concurrent examination of executive function difficulties and relevant ADHD symptoms within a college population. Future work could further examine the effect of sex on MIMIC models of ADHD to determine if sex moderates associations of EF deficits and ADHD.

Overall, findings provide support for the validity of the BDEFS as a tool for examining EF in college students with a few stipulations. Sex, inattention, and hyperactivity-impulsivity symptoms appeared to moderately impact performance of some items and were also clearly related to the BDEFS factor means. Future work may benefit from investigating sex differences more broadly across these EF factors within young adult college students to determine if these differences are related to other validity components of the BDEFS. Findings also suggested that ADHD symptoms impacted item performance and mean levels of each of the five BDEFS factors. Thus, considering ADHD symptomatology and diagnostic history may be particularly relevant when assessing EF in college students. Despite this, results suggest that the BDEFS is a robust measure to capture multiple component EF deficits among college students.

Limitations

The present study should be evaluated in light of a few limitations. For example, a portion of the larger sample could not be included in analyses due to the non-random missing data. However, the sample utilized remained relatively large, with a total of 1,311 participants.

Results should also be considered in light of the different possibilities that exist for interpreting models and making decisions regarding ESEM cross-loadings. Specifically, one of the advantages of ESEM is that, compared to CFA, it can more accurately model the reality of cross-loadings. In this study, we essentially considered cross-loadings to be

"problematic". While this approach was desirable to simplify the structure of the BDEFS (i.e., items load on one and only one factor), it may not always be the optimal approach. For example, this approach may involve removal of some items that might be excellent indicators of a certain construct (i.e., factor) because they also load on a second related factor. Again, although the approach to eliminate items with high cross-loadings seems justified for our purposes, it is important to acknowledge the trade-off in doing so. Additionally, some of the "problematic" items may reflect that they are double barreled questions (e.g., "cannot focus my attention on tasks or work as well as others"), which could have impacted the cross-loadings.

Moreover, our study relied on self-report of EF as well as ADHD symptoms, which may bias results due to common method variance. Additionally, sole reliance of self-report of ADHD symptoms may be problematic in adults, as adults with ADHD may not always have accurate insights regarding their own behavior (e.g., underreporting the severity of their symptoms) (Sandra Kooij et al., 2008). That said, it is important to note that the focus of this research was to validate the BDEFS measure in a *non-selected* sample of college students, so the study is strengthened by not recruiting for diagnosis-specific, or treatment-seeking individuals. Further, because the focus of this study was on validation in a college sample, the age range was substantially narrower than the BDEFS normative sample (18–28 years vs. 18–81 years). It may be useful if future work replicates the invariance analyses of the present study in a sample with a larger age range. Future research should also explore if these findings apply to *non-college* students of similar age. Finally, we did not evaluate other forms of validity (convergent, discriminant, and predictive validity). Future work should continue to evaluate the validity of the BDEFS measure in college populations to determine its utility in predicting relevant academic performance outcomes (e.g., Grade Point Average, college attrition/graduation).

Conclusion and Implications

Our findings suggest that the five-factor model of the BDEFS is supported in an English-speaking college population and that its scores reflecting EF deficits across five domains can be measured with adequate reliability. Further, our findings provide support for the validity of interpretations of those scores across age, sex, and ADHD symptomatology. There was general support for invariance across sex and age. The present study contributes to the field by evaluating the BDEFS in a non-clinical, multi-site college population, where identifying EF deficits could have important implications.

Support for the validity of the BDEFS indicates that this measure may be a useful tool for identifying at-risk college students. For example, college students with EF deficits report significantly higher rates of nonmedical use of stimulants compared to those without EF deficits (Munro et al., 2017), and have difficulties with evaluating the consequences of their actions (Pharo, Sim, Graham, Gross, & Hayne, 2011), even if they do not have a clinical disorder that predisposes them to increased risk-taking (Romer, 2010). Evaluation of EF deficits in college students could assist with identifying those at risk for such problems, targeting practical coping tools, and helping create and attain appropriate career goals. Further, most colleges are equipped with career planning centers, in which students work

with a career advisor to make their future professional plans. In this context, the BDEFS could be used to identify areas of impairment and consequently used to implement more effective career counseling with students (Prevatt & Yelland, 2015). For example, it can be speculated that in the future, the BDEFS may help a career counselor to make initial hypotheses about a student's strengths and weaknesses and appropriately guide them in choosing classes, setting goals, and learning strategies that build on their strengths to meet those goals. Additionally, the BDEFS could be used in college settings to identify students who seem to be having so much difficulty that they need more formal testing to determine if there are associated clinical problems, such as ADHD. Taken together, it appears this measure's utility may extend to multiple populations, and can be used as an efficient way to identify students' EF difficulties in order to provide better career guidance in the service of helping them achieve success.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References

Barkley RA (1997). ADHD and the Nature of Self-control: Guilford Press.

Barkley RA (2010). Differential diagnosis of adults with ADHD: the role of executive function and self-regulation. J Clin Psychiatry, 71(7), e17. doi:10.4088/JCP.9066tx1c [PubMed: 20667287]

Barkley RA (2011). Barkley deficits in executive functioning scale (BDEFS for adults). New York: Guilford Press.

Barkley RA (2012). Barkley Deficits in Executive Functioning Scale--Children and Adolescents (BDEFS-CA): Guilford Press.

Barkley RA, & Fischer M (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: self-reported executive function (EF) deficits versus EF tests. Developmental Neuropsychology, 36(2), 137–161. doi:10.1080/87565641.2010.549877 [PubMed: 21347918]

Barkley RA, & Murphy KR (2011). The Nature of Executive Function (EF) Deficits in Daily Life Activities in Adults with ADHD and Their Relationship to Performance on EF Tests. Journal of Psychopathology and Behavioral Assessment, 33(2), 137–158.

Becker SP, Jarrett MA, Luebbe AM, Garner AA, Burns GL, & Kofler MJ (2018). Sleep in a large, multi-university sample of college students: sleep problem prevalence, sex differences, and mental health correlates. Sleep Health, 4(2), 174–181. [PubMed: 29555131]

Biederman J, Petty C, Fried R, Fontanella J, Doyle AE, Seidman LJ, & Faraone SV (2006). Impact of psychometrically defined deficits of executive functioning in adults with attention deficit hyperactivity disorder. American Journal of Psychiatry, 163(10), 1730–1738. doi:10.1176/ajp.2006.163.10.1730

Bogod NM, Mateer CA, & MacDonald SW (2003). Self-awareness after traumatic brain injury: a comparison of measures and their relationship to executive functions. Journal of International Neuropsychological Society, 9(3), 450–458. doi:10.1017/S1355617703930104

Booth T, & Hughes DJ (2014). Exploratory structural equation modeling of personality data. Assessment, 21(3), 260–271. doi:10.1177/1073191114528029 [PubMed: 24682748]

Browne MW, & Cudeck R (1993). Alternative ways of assessing model fit. Bollen KA & Long JS (Eds.),Testing structural equation models (pp. 136–162). London, England: Sage.

Burgess PW, Alderman N, Evans J, Emslie H, & Wilson BA (1998). The ecological validity of tests of executive function. Journal of International Neuropsychological Society, 4(6), 547–558.

Burns LR, Dittman K, Nguyen N, & Mitchelson JK (2000). Academic procrastination, perfectionism, and control: Associations with vigilant and avoidant coping. Journal of Social Behavior and Personality, 5, 35–46.

Carragher N, Krueger RF, Eaton NR, Markon KE, Keyes KM, Blanco C, … Hasin DS (2014). ADHD and the externalizing spectrum: direct comparison of categorical, continuous, and hybrid models of liability in a nationally representative sample. Social Psychiatry and Psychiatry Epidemiology, 49(8), 1307–1317. doi:10.1007/s00127-013-0770-3

Cole DA, Perkins CE, & Zelkowitz RL (2016). Impact of homogeneous and heterogeneous parceling strategies when latent variables represent multidimensional constructs. Psychological Methods, 21(2), 164–174. doi:10.1037/met0000047 [PubMed: 26323000]

Costello AB, & Osborne JW (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research & evaluation, 10(7), 1–9.

Dehili VM, Prevatt F, & Coffman TP (2017). An Analysis of the Barkley Deficits in Executive Functioning Scale in a College Population: Does It Predict Symptoms of ADHD Better Than a Visual-Search Task? Journal of Attention Disorders, 21(7), 567–574. doi:10.1177/1087054713498932 [PubMed: 24026813]

Friedman NP, Miyake A, Altamirano LJ, Corley RP, Young SE, Rhea SA, & Hewitt JK (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. Developmental Psychology, 52(2), 326–340. doi:10.1037/dev0000075 [PubMed: 26619323]

Friedman NP, Miyake A, Young SE, Defries JC, Corley RP, & Hewitt JK (2008). Individual differences in executive functions are almost entirely genetic in origin. Journal of Experimental Psychology General, 137(2), 201–225. doi:10.1037/0096-3445.137.2.201 [PubMed: 18473654]

Garrido LE, Abad FJ, & Ponsoda V (2013). A new look at Horn's parallel analysis with ordinal variables. Psychological Methods, 18(4), 454–474. doi:10.1037/a0030005 [PubMed: 23046000]

Gerstorf D, Siedlecki KL, Tucker-Drob EM, & Salthouse TA (2008). Executive dysfunctions across adulthood: measurement properties and correlates of the DEX self-report questionnaire. Aging, Neuropsychology, and Cognition, 15(4), 424–445. doi:10.1080/13825580701640374

Goldstein S, & Naglieri JA (2014). Handbook of Executive Functioning. New York: Springer.

Hu LT, & Bentler PM (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model mis- specification. Psychological Methods, 3, 424–453.

Hu LT, & Bentler PM (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55.

Huizinga M, Dolan CV, & van der Molen MW (2006). Age-related change in executive function: developmental trends and a latent variable analysis. Neuropsychologia, 44(11), 2017–2036. doi:10.1016/j.neuropsychologia.2006.01.010 [PubMed: 16527316]

Janssen GT, De Mey HR, & Egger JI (2009). Executive functioning in college students: evaluation of the Dutch executive function index (EFI-NL). International Journal of Neuroscience, 119(6), 792–805. doi:10.1080/00207450802333979

Jarrett MA (2016). Attention-deficit/hyperactivity disorder (ADHD) symptoms, anxiety symptoms, and executive functioning in emerging adults. Psychological Assessment, 28(2), 245–250. doi:10.1037/pas0000190 [PubMed: 26121381]

Kachgal MM, Hansen SL, & Nutter KJ (2001). Academic procrastination/intervention: Strategies and recommendations. Journal of Developmental Education, 25, 14–24.

Kamradt JM, Ullsperger JM, & Nikolas MA (2014). Executive function assessment and adult attention-deficit/hyperactivity disorder: tasks versus ratings on the Barkley deficits in executive functioning scale. Psychological Assessment, 26(4), 1095–1105. doi:10.1037/pas0000006 [PubMed: 24885846]

LaCount PA, Hartung CM, Shelton CR, & Stevens AE (2018). Efficacy of an Organizational Skills Intervention for College Students With ADHD Symptomatology and Academic Difficulties. Journal of Attention Disorders, 22(4), 356–367. doi:10.1177/1087054715594423 [PubMed: 26253149]

Larsson H, Anckarsater H, Rastam M, Chang Z, & Lichtenstein P (2012). Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. Journal of Child Psychology and Psychiatry, 53(1), 73–80. doi:10.1111/j.1469-7610.2011.02467.x [PubMed: 21923806]

Lezak MD (1995). Neuropsychological Assessment. New York: Oxford University Press.

Little TD, Rhemtulla M, Gibson K, & Schoemann AM (2013). Why the items versus parcels controversy needn't be one. Psychological Methods, 18(3), 285–300. doi:10.1037/a0033266 [PubMed: 23834418]

Marcus DK, Norris AL, & Coccaro EF (2012). The latent structure of attention deficit/hyperactivity disorder in an adult sample. Journal of Psychiatry Research, 46(6), 782–789. doi:10.1016/j.jpsychires.2012.03.010

Marsh HW, Ludtke O, Nagengast B, Morin AJ, & Von Davier M (2013). Why item parcels are (almost) never appropriate: two wrongs do not make a right--camouflaging misspecification with item parcels in CFA models. Psychological Methods, 18(3), 257–284. doi:10.1037/a0032773 [PubMed: 23834417]

Mischel W, Ayduk O, Berman MG, Casey BJ, Gotlib IH, Jonides J, … Shoda Y (2011). 'Willpower' over the life span: decomposing self-regulation. Social Cognitive and Affective Neuroscience, 6(2), 252–256. doi:10.1093/scan/nsq081 [PubMed: 20855294]

Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington H, … Caspi A (2011). A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences of the United States of America, 108(7), 2693–2698. doi:10.1073/pnas.1010076108 [PubMed: 21262822]

Moon SM, & Illingworth AJ (2005). Exploring the dynamic nature of procrastination: A latent growth curve analysis of academic procrastination. Personality and Individual Differences, 38, 297–309.

Morisano D, Hirsh JB, Peterson JB, Pihl RO, & Shore BM (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. Journal of Applied Psychology, 95(2), 255–264. doi:10.1037/a0018478

Munro BA, Weyandt LL, Marraccini ME, & Oster DR (2017). The relationship between nonmedical use of prescription stimulants, executive functioning and academic outcomes. Addictive Behavior, 65, 250–257. doi:10.1016/j.addbeh.2016.08.023

Muthen LK, & Muthen BO (2011). Mplus user's guide, 6th edition. Los Angeles: CA: Author.

Naglieri JA, & Goldstein S (2017). Comprehensive Executive Function Inventory Adult – (CEFI Adult): MHS.

O'Hearn K, Asato M, Ordaz S, & Luna B (2008). Neurodevelopment and executive function in autism. Dev Psychopathol, 20(4), 1103–1132. doi:10.1017/S0954579408000527 [PubMed: 18838033]

Onwuegbuzie AJ (2004). Academic procrastination and statistics anxiety. Assessment and Evaluation in Higher Education, 29, 3–19.

Oppenheimer DM, Meyvis T, & Davidenko N (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45(4), 867–872.

Pennington BF, & Ozonoff S (1996). Executive functions and developmental psychopathology. Journal of Child Psychology and Psychiatry, 37(1), 51–87. [PubMed: 8655658]

Petersen R, Lavelle E, & Guarino AJ (2006). The relationship between college students' executive functioning and study strategies. Journal of College Reading and Learning, 36(2), 59–67.

Pharo H, Sim C, Graham M, Gross J, & Hayne H (2011). Risky business: executive function, personality, and reckless behavior during adolescence and emerging adulthood. Behavioral Neuroscience, 125(6), 970–978. doi:10.1037/a0025768 [PubMed: 22004262]

Prevatt F, & Yelland S (2015). An Empirical Evaluation of ADHD Coaching in College Students. J Atten Disord, 19(8), 666–677. doi:10.1177/1087054713480036 [PubMed: 23509112]

Rabin LA, Fogel J, & Nutter-Upham KE (2011). Academic procrastination in college students: the role of self-reported executive function. Journal of Clinical and Experimental Neuropsychologyl, 33(3), 344–357. doi:10.1080/13803395.2010.518597

Romer D (2010). Adolescent risk taking, impulsivity, and brain development: implications for prevention. Developmental Psychobiology, 52(3), 263–276. doi:10.1002/dev.20442 [PubMed: 20175097]

Roth RM, Isquith PK, & Gioia GA (2005). BRIEF-A: Behavior Rating Inventory of Executive Function—Adult Version manual. Odessa, FL: Psychological Assessment Resources.

Sandra Kooij JJ, Marije Boonstra A, Swinkels SH, Bekker EM, de Noord I, & Buitelaar JK (2008). Reliability, validity, and utility of instruments for self-report and informant report concerning symptoms of ADHD in adult patients. Journal of Attention Disorders, 11(4), 445–458. doi:10.1177/1087054707299367 [PubMed: 18083961]

Schutz PA, White VE, & Lanehart SL (2000). Core goals and their relationship to semester sub goals and academic performance. Journal of College Student Retention, 2, 13–28.

Snyder HR (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychological Bulletin, 139(1), 81–132. doi:10.1037/a0028727 [PubMed: 22642228]

Spinella M (2005). Self-rated executive function: development of the executive function index. International Journal of Neuroscience, 115(5), 649–667. doi:10.1080/00207450590524304

Tice DM, & Baumeister RF (1997). Longitudinal study of procrastination, performance, stress and health: The costs and benefits of dawdling. Psychological Science, 8, 454–458.

Timmerman ME, & Lorenzo-Seva U (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods, 16(2), 209–220. doi:10.1037/a0023353 [PubMed: 21500916]

Toplak ME, West RF, & Stanovich KE (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. doi:10.1111/jcpp.12001 [PubMed: 23057693]

U.S. Department of Education, National Center for Education Statistics. (2018). Digest of Education Statistics, 2016 (NCES 2017–094), Chapter 3.

Velez-Pastrana MC, Gonzalez RA, Rodriguez Cardona J, Purcell Baerga P, Alicea Rodriguez A, & Levin FR (2016). Psychometric properties of the Barkley Deficits in Executive Functioning Scale: A Spanish-Language Version in a community sample of puerto rican adults. Psychological Assessment, 28(5), 483–498. doi:10.1037/pas0000171 [PubMed: 26302104]

Weyandt LL (2005). Executive function in children, adolescents, and adults with attention deficit hyperactivity disorder: introduction to the special issue. Developmental Neuropsychologyl, 27(1), 1–10. doi:10.1207/s15326942dn2701_1

Willcutt EG, Doyle AE, Nigg JT, Faraone SV, & Pennington BF (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biological Psychiatry, 57(11), 1336–1346. doi:10.1016/j.biopsych.2005.02.006 [PubMed: 15950006]

Wodka EL, Mostofsky SH, Prahme C, Gidley Larson JC, Loftis C, Denckla MB, & Mahone EM (2008). Process examination of executive function in ADHD: sex and subtype effects. Clinical Neuropsychology, 22(5), 826–841. doi:10.1080/13854040701563583

Table 1.

Demographic Characteristics and Descriptive Statistics of US College Students (N=1,311)

	%	N
Race		
White	79.8	1,046
Asian/Asian American	7.5	98
Black/African American	6.6	87
Native Hawaiian/Other Pacific Islander	0.4	5
merican Indian/Alaska Native	0.6	8
Biracial/Multiracial	4.6	60
Ethnicity		
Hispanic/Latino	6.0	79
Non Hispanic/Latino	94.0	1,229
Mother Education		
Some high school	2.3	30
Completed high school	11.3	148
Some college	16.0	210
Completed college	47.3	620
Completed graduate school	23.0	302
Father Education		
Some high school	3.2	42
Completed high school	12.8	168
Some college	14.1	185
Completed college	39.6	519
Completed graduate school	30.1	395
Parental Income		
Up to $39,999	6.8	89
$40,000–$79,999	14.4	188
$80,000–$119,999	17.7	232
$120,000–$159,999	14.0	183
$160,000–$199,999	8.0	105
$200,000 or more	18.5	242
Year in College		
1st Year	64.1	840
2nd Year	18.8	247
3rd Year	10.7	140
4th Year	5.9	77
Other	0.5	7
Self-Report of Professional Diagnosis		
No diagnoses	74.7	979
Anxiety	13.5	177
Depression	13.0	171

	%	N
ADD/ADHD	9.2	121
Learning Disorder	2.4	31
Personality Disorder	0.5	7
Bipolar Disorder	0.7	9
Alcohol abuse/disorder	0.2	3
Autism/Asperger's	0.1	1
Schizophrenia	0.0	0

Note. Participants could select multiple diagnoses.

Table 2.

Fit Indices of ESEM vs. CFA of Barkley Deficits in Executive Function Scale, CFA in Full Sample All Items vs. Subsets Excluding Items with High Cross-Loadings, and CFA in Invariance Subgroups.

Model	χ^2 (df)	RMSEA	RMSEA 90% CI	CFI	TLI
ESEM vs. CFA					
ESEM (n=656)	6290.700 (3481)*	0.035	[0.034, 0.036]	0.960	0.960
CFA (n=655)	7826.400 (3817)*	0.040	[0.039, 0.041]	0.940	0.940
CFA All Items vs. Subsets					
All items	14163.100 (3817)*	0.045	[0.045, 0.046]	0.930	0.920
>.25	8731.580 (2267)*	0.047	[0.046, 0.048]	0.940	0.940
>.30	11322.436 (3070)*	0.045	[0.044, 0.046]	0.930	0.930
>.35	13396.330 (3644)*	0.045	[0.044, 0.046]	0.930	0.930
CFA Group Analyses					
Males	6430.100 (3817)*	0.039	[0.037, 0.040]	0.950	0.950
Females	9845.540 (3817) *	0.043	[0.042, 0.044]	0.930	0.930
Younger (19)	10359.590 (3817)	0.043	[0.042, 0.044]	0.933	0.932
Older (20)	6242.100 (3817)*	0.041	[0.039, 0.042]	0.940	0.939
Low Inattention	7965.630 (3817)*	0.038	[0.036, 0.039]	0.937	0.935
High Inattention	8237.730 (3817)*	0.040	[0.039, 0.041]	0.939	0.938
Low Hyperactivity	8124.530 (3817)*	0.038	[0.036, 0.040]	0.943	0.941
High Hyperactivity	8466.440 (3817)*	0.041	[0.038, 0.044]	0.938	0.935

Note. Estimations based on weighted least square-mean and variance. RMSEA root-mean-square error of approximation; CI confidence interval; CFI comparative fit index; TFI Tucker–Lewis Index. *ESEM vs. CFA*: CFA (confirmatory factor analysis) conducted in the first half of the sample (n=655) and ESEM (exploratory structural equation model) conducted in the second half of the sample (n=656). *CFA all items vs. subsets and CFA Group Analyses* presented were conducted within the full sample (n=1,311).

Table 3.

Model Fit Statistics for CFA Invariance Tests Across Sex Full Sample of US college students (n=1,311).

Model	χ^2DIFF (DF)	CFI	CFI	RMSEA	95% CI
All Items					
Equal Forms	--------------	.940	------	.041	.040, .042
Full Invariance	577.120 (351)[*]	.953	−.013	.035	.034, .036
Partial Invariance	452.290 (313)[*]	.954	−.014	.035	.034, .036
Reduced BDEFS					
Equal Forms	--------------	.959	------	.046	.045, .048
Full Invariance	216.710 (186)	.969	−.010	.039	.037, .040

Note.

[*] $p<.001$.

χ2DIFF= Chi-Square Test for Difference Testing for WLSMV estimation, CFI=Comparative Fit Index, CFI=change in CFI, RMSEA= Root Mean Square Error of Approximation, 95% CI= 95% confidence interval for RMSEA. For all model evaluations, full invariance and partial invariance models compared to equal forms models based on chi-square DIFFTEST for WLSMV estimation and the change in CFI.

Table 4.

Confirmatory Factor Analysis Multiple Indicator Multiple Causes Model Coefficients for BDEFS Factor Differences by Sex, Age, and Attention-Deficit/ Hyperactivity Disorder Symptom Dimensions.

Covariate	RMSEA	F1	F2	F3	F4	F5	Direct Effects	Items
Sex	0.064	−.02	.11	−.15**	−.17**	.23**	Yes	46, 61, 64, 78
Age	0.064	−.01	−.04	−.06*	−.05	−.06*	No	NA
Inattention	0.060	.34**	.30**	.25**	.26**	.21**	Yes	1, 22, 39, 41, 43, 65
Hyperactivity-Impulsivity	0.062	.34**	.36**	.39**	.26**	.33**	Yes	1, 22, 39, 79

Note. RMSEA=root mean square error of approximation. *F1*= Self-Management to Time, *F2*=Self-Organization, *F3*=Self-Restraint, *F4*=Self-Motivation, *F5*=Self=Regulation of Emotions. *Items*= items that were significantly associated with the corresponding covariates.

*
p<.05,

**
p<.01.