School Mental Health (2017) 9:143–156
DOI 10.1007/s12310-016-9205-0

ORIGINAL PAPER

Evaluating the Factor Validity of the Children's Organizational Skills Scale in Youth with ADHD

Stephen J. Molitor[1] · **Joshua M. Langberg**[1] · **Steven W. Evans**[2] · **Melissa R. Dvorsky**[1] · **Elizaveta Bourchtein**[1] · **Laura D. Eddy**[1] · **Zoe R. Smith**[1] · **Lauren E. Oddo**[1]

Published online: 27 December 2016

Abstract Children and adolescents with ADHD often have difficulties with organization, time management, and planning skills, and these skills are a common target of intervention. A limited array of tools for measuring these abilities in youth is available, and one of the most prominent measures is the Children's Organizational Skills Scale (COSS). Although the COSS fills an important need, a replication of the COSS factor structure outside of initial measure development has not been conducted in any population. Given that the COSS is frequently used in ADHD research, the current study evaluated the factor structure of the parent-rated COSS in a sample ($N = 619$) of adolescents with ADHD. Results indicated that the original factor structure could be replicated, although the use of item parcels appeared to affect model fit statistics. An alternative bi-factor model was also tested that did not require the use of parcels, with results suggesting similar model fit in comparison with the original factor structure. Exploratory validity tests indicated that the domain-general factor of the bi-factor model appears related to broad executive functioning abilities.

Keywords ADHD · Organizational skills · Academic functioning · Adolescence

✉ Stephen J. Molitor
molitorsj@mymail.vcu.edu

[1] Virginia Commonwealth University, 806 W. Franklin Street, P.O. Box 842018, Richmond, VA 23284-2018, USA

[2] Ohio University, 200 Porter Hall, Athens, OH 45701, USA

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common childhood disorders with prevalence estimates around 7% of school-age children (Visser et al., 2013). Children and adolescents with ADHD frequently experience clinically significant academic impairment (DuPaul & Langberg, 2014; Frazier, Youngstrom, Glutting, & Watkins, 2007), and difficulty at school is one of the most prominent reasons youth with ADHD are referred for treatment (Loe & Feldman, 2007; Wolraich et al., 2005). The academic impairments of students with ADHD tend to persist or increase in severity throughout their educational careers and into post-secondary education (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2007; Frazier et al., 2007; Weyandt et al., 2013). Given the chronic nature of academic impairment in individuals with ADHD, it is important to understand the mechanisms that lead to these poor academic outcomes.

The main symptoms of the disorder (i.e., inattention, hyperactivity, and impulsivity) can directly interfere with learning by making it difficult to sustain attention during instruction time and/or disrupting the classroom environment (Zoromski, Owens, Evans & Brady, 2015). Other related deficits can also significantly impact academic functioning. For example, youth with ADHD frequently experience deficits in organization, time management, and planning skills (i.e., OTMP skills) (Barkley, 1997; Harrier & DeOrnellas, 2005). Although only one core symptom from the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) ADHD criteria explicitly addresses these skills (i.e., difficulty organizing oneself), several other core symptoms are closely related (e.g., difficulties with details, fails to finish tasks, losing things, forgetting deadlines/appointments) (American Psychiatric

Association, 2013). These symptoms manifest in many ways in school, including difficulty managing physical materials necessary to complete daily tasks (e.g., school supplies, homework, clothes) and problems managing and organizing time (e.g., completing assignments by the due date, accurately estimating time needed to complete work, and planning ahead for examinations and long-term projects) (Booster, DuPaul, Eiraldi, & Power, 2010; Gureasko-Moore, DuPaul, & White, 2007; Langberg et al., 2011). Empirical studies document that OTMP skills are often impaired in youth with ADHD and are associated with academic outcomes, both cross-sectionally and longitudinally, above and beyond the impact of ADHD symptoms (Evans et al., 2009; Langberg, Dvorsky & Evans, 2013; Langberg et al., 2011).

Despite the importance of OTMP skills for successful academic functioning, few methods are available to specifically measure these abilities. Neuropsychological tasks have been developed that are hypothesized to tap some aspects of OTMP skills (e.g., Tower of London to measure planning abilities; Shallice, 1982), but questions have been raised about whether these tasks accurately capture everyday, real-world, OTMP behaviors (Toplak, West, & Stanovich, 2013). Rating scales have also been developed that measure these skills indirectly. For example, the Homework Problems Checklist (HPC; Anesko, Schoiock, Ramirez & Levine, 1987) is a parent-completed measure that contains items that ask how frequently a child forgets an assignment or does not have all materials necessary to complete an assignment. The Homework Performance Questionnaire (HPQ; Power, Dombrowski, Watkins, Mautone, & Eagle, 2007; Power et al., 2015) also provides an assessment of homework performance in terms of competence and skills and has both parent and teacher versions. However, items from both measures are specific to homework problems and do not measure many aspects of OTMP abilities. Another measure that contains OTMP items/subscales is the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy & Kenworthy, 2000a, b). In particular, two of the subscales, Plan/Organize and Organization of Materials, directly address OTMP behaviors and include questions about academic impairment related to OTMP. However, the Children's Organizational Skills Scale (COSS; Abikoff & Gallagher, 2009) is the only measure focused strictly on evaluating OTMP skills and offers insights into the planning and organizational skills of youth as applied broadly in daily life. The scale was developed using a large, nationally representative sample ($N = 1440$; ages 8–13), which included the intentional collection of a subsample of children diagnosed with ADHD ($n = 338$) given the high prevalence of OTMP deficits in this population.

During the development of the parent and teacher versions of the COSS, an exploratory factor analysis (EFA) elicited a three-factor structure that used a subset of the items on the COSS (28 of the 58 total items). The factors were termed Task Planning, Organized Actions, and Memory/Materials Management based upon the items within each factor. The Task Planning factor is thought to measure the use of skills to prepare a task for completion, such as breaking a large task down into parts. The Organized Actions factor is thought to evaluate how frequently youth use tools (e.g., calendars, to-do lists) or routines (e.g., making an essay outline before writing) to aid task completion. The Memory/Materials Management factor is designed to measure how well youth can keep track of relevant task information (e.g., homework assignments, due dates) as well as the physical materials (e.g., books, worksheets, pens/pencils) needed to complete tasks. The remaining 30 items contribute only to the Total Score generated by the COSS; some of these items are part of the Inconsistency Index to detect invalid responding, whereas other items were retained because they "… pertained to behaviors or outcomes believed to be clinically important" (p. 48, Abikoff & Gallagher, 2009). The technical manual encourages the interpretation of both the Total Score and each of the scales elicited from the factor analysis. A confirmatory factor analysis (CFA) was then conducted, and two of the items were removed from the parent report form because parents would not likely have knowledge about the items (e.g., "The child's desk is neat"). The final 26-item factor structure was determined to fit the data well (Abikoff & Gallagher, 2009). Given the established difficulties with OTMP skills that children and adolescents with ADHD experience, it is not surprising that the COSS is heavily used in research focused on these populations. For example, the COSS is frequently used as an outcome measure for ADHD intervention studies, including pharmacological interventions (e.g., Abikoff et al., 2009) and psychosocial interventions (e.g., Evans et al., 2016; Langberg, Epstein, Becker, Girio-Herrera, & Vaughn, 2012; Pfiffner, Villodas, Kaiser, Rooney, & McBurnett, 2013).

Although the COSS fills an important gap in the study of OTMP skills in youth with ADHD, it is a relatively new measure that has received no evaluation of factor structure other than what is reported in the technical manual. Further, there are some limitations associated with the factor analysis procedures reported in the technical manual. First, several decision points were not described in detail, such as exactly how eigenvalues, scree plots, and factor interpretability were used to elicit the optimal number of factors, which is inconsistent with best-practice recommendations (Henson & Roberts, 2006) and makes it difficult for readers to understand the developers' justification for the final factors. Second, the developers appeared to contradict some of their established procedures. For example, the manual reported that item cross-loadings of 0.40 would not be retained in the final factor structure. However, two retained items violated this cross-loading threshold ("hard to find papers" and "forgets about tests"). Third, it is unclear how the 30 items not included in their analyses load onto the established factors or onto other, unreported factors. This makes it more difficult to understand the relationships between all items found on the measure.

Another, more problematic limitation of the reported CFA results was the use of item parcels. Item parceling is a statistical procedure sometimes used in factor analysis as an alternative to traditional CFA procedures. When item parcels are used, a sum or average of two or more items are calculated and that value is used as a single observed indicator of the factor structure instead of each of the individual item values. Parceling reduces the total number of indicators in a model, which can help ensure that a model converges through reduced parameter estimates and lower indicator/sample size ratios as long as all necessary assumptions are met (Little, Rhemtulla, Gibson, & Schoemann, 2013; Matsunaga, 2008). Opponents of the process argue that in practice, few researchers provide the necessary theoretical rationale for parceling items or check that all necessary assumptions are met (Bandalos & Finney, 2001). Additionally, there are concerns that parceling items may artificially inflate fit indices in comparison with item-level CFAs because they can reduce the effects of cross-loading items and items with correlated measurement errors, which generally lower fit index values (Bandalos, 2002; Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013). No factor structure analyses were reported without item parceling in the COSS technical manual and a rationale for clustering was not provided, meaning that a further examination of the potential effect of item parceling on the overall fit of the factor structure is warranted.

If item parceling does indeed inflate the fit indices when confirming the factor structure of the COSS, two important follow-up questions should be posed. First, it should be determined whether acceptable fit indices are still achieved when confirming the factor structure without parceling items. If acceptable fit cannot be achieved without item parcels with the current factor structure, then an alternative factor structure would need to be explored. Deciding which specific structures to explore, however, requires an evaluation of other relevant constructs related to OTMP skills. OTMP skills are often conceptualized as one component of the broader cognitive system of executive functions (EF). EF broadly refers to the cognitive abilities needed to successfully engage in goal-directed behavior, both within a given situation and over time (Pennington & Ozonoff, 1996). These abilities are often divided into two distinct subdomains (Prencipe et al., 2011). The first subdomain is hot EF, which is responsible for short-term, emotionally driven behavior (e.g., inhibition, emotion regulation). Cool EF, in contrast, is responsible for long-term planning and self-regulation behaviors; OTMP skills fall under this subdomain. Importantly, many tasks require the coordination of multiple EF abilities, making it difficult to untangle the influence of one ability (e.g., planning) from another (e.g., execution) (Hughes & Graham, 2002). Given this connection between OTMP skills and other EF abilities, it is possible that a rater's responses measure not only OTMP skills specifically, but also broader EF abilities. Therefore, a more complex factor structure that attempts to account for these relationships may provide an alternative solution. Finding an alternative factor structure may be particularly relevant when using the COSS with youth with ADHD as they tend to exhibit poorer EF abilities in comparison with others without the disorder, especially in relation to organization and planning (Harrier & DeOrnellas, 2005; Shimoni, Engel-Yeger, & Tirosh, 2012).

One alternative approach—a bi-factor model—captures variance in responses to a particular item on a measure that may be accounted for by a broad construct (i.e., a domain-general construct) as well as a narrower construct (i.e., a domain-specific construct) (for a review of bi-factor measurement modeling, see Reise, 2012). In the context of the COSS, the original factors may represent domain-specific constructs. However, several of the items presented in the COSS may tap more than one construct, and its variance therefore cannot be adequately accounted for by a specific factor. For example, the COSS contains the items "My child is organized" and "My child gets stressed out or overwhelmed trying to get his/her work organized." Using only these prompts to guide their ratings, it would be difficult to know whether raters assume these items refer to a single domain or to a broader set of abilities that contribute as a whole to the perception of a child being "organized." The variance observed in these items may therefore be better represented by both a specific latent factor and a general latent factor.

Current Study

Given the importance of OTMP skills for academic success and the widespread use of the COSS in ADHD intervention work, in the current study we sought to evaluate the factor structure that emerged during the development of the parent-rated COSS and to evaluate an alternative, theoretically supported structure. The study had two specific aims. First, the structure proposed by the measure developers was evaluated using CFA with a sample of middle school students ($N = 619$) diagnosed with ADHD. The CFA was conducted both with item parceling (as used in the original CFA) and without, in order to evaluate whether item parceling significantly inflated the fit statistics of the

CFA. The second aim was to evaluate an alternative bi-factor model based upon the theory that variance in responses to COSS items is affected by both the domain-specific factors and a domain-general factor. The proposed bi-factor model was compared to the original model outlined in the COSS manual and the three-factor model without item parcels. A third, more exploratory aim was to preliminarily evaluate the construct validity of the domain-general factor as associated with several scale scores from the BRIEF. Given that some of the OTMP items on the BRIEF overlap with items on the COSS, it was hypothesized that the domain-general factor would be significantly correlated with total BRIEF scores as well as Metacognition Index scores. It was also hypothesized that the domain-general factor would be more strongly correlated with subscales on the BRIEF directly related to OTMP skills (e.g., Plan/Organize, Organization of Materials) than other subscales.

Methods

Participants

To conduct this study with a large sample of middle school adolescents comprehensively diagnosed with ADHD, participants from three separate studies were combined into a single dataset. Importantly, the recruitment and inclusion/exclusion procedures were identical across the three studies. The focus in the present study is on COSS parent ratings for two reasons. First, parent but not teacher ratings were collected consistently across all three studies during the baseline evaluations (near 100% measurement completion rate for parent ratings). Second, given that participants were all in middle school, focusing on parent ratings circumvents complex measurement issues that often arise when obtaining the teacher perspective in middle and high schools, such as low response correlations between teachers and concerns that teachers may not be able to reliably observe and rate some OTMP behaviors (Evans, Allen, Moore, & Strauss, 2005; Narad et al., 2015). All participants were recruited as part of studies investigating the efficacy of school-based behavior interventions for middle school students with ADHD. Students and their families were recruited from fifteen rural, urban, and suburban middle schools in the eastern and midwestern USA. All data used in this study were collected during an initial assessment visit before any intervention was administered. A total of 619 adolescents formally diagnosed with ADHD were included in the study. Table 1 provides information regarding the demographic characteristics of each sample.

Recruitment and Evaluation Procedure

All three studies were reviewed and approved by the appropriate institutional review board (IRB), and all participants consented to participate in the study. All participants were recruited through study announcement letters sent to all parents at participating middle schools, posted fliers, and direct referrals from school staff. No prior diagnosis of ADHD was required for participation in the study. If the parents or caretakers of a potential participant were interested in the study, they contacted the study staff, who conducted a brief telephone screen. The phone screens were identical across the three studies, and parents were asked whether their child had previously received a formal ADHD diagnosis. If parents did not endorse a prior ADHD diagnosis, at least 4 of 9 inattentive symptoms had to be verbally endorsed as often or very often for the inclusion/exclusion assessment to be scheduled.

During the inclusion/exclusion evaluation, students were comprehensively assessed for an ADHD diagnosis. Adolescents were considered eligible for study participation if they met full DSM-IV-TR diagnostic criteria for ADHD—Predominantly Inattentive or Combined subtypes on the parent version of the Children's Interview for Psychiatric Syndromes (ChIPS; Weller, Weller, Fristad, Rooney, & Schecter, 2000), experienced significant impairment due to ADHD symptoms according to parent and teacher ratings, had an estimated FSIQ of at least 80 according to performance on the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003), and did not meet diagnostic criteria for any pervasive developmental disorder, bipolar disorder, psychosis, or obsessive–compulsive disorder. The ChIPS is a structured interview designed to evaluate the mental health of children and adolescents ages 6–18 that has demonstrated good construct validity in relation to other semi-structured interviews and good inter-rater agreement (Fristad, Teare, Weller, Weller, & Salmon, 1998; Leffler, Riebel, & Hughes, 2014).

The Children's Organizational Skills Scale: Parent Report (Abikoff & Gallagher, 2009)

The COSS parent is a rating scale composed of 58 items designed to assess the frequency of behaviors associated with organization and planning in both the home and school environment and 8 items assessing how much a child's current OTMP abilities are impairing their functioning. Items are rated using a 4-point rating scale (1 = "Hardly ever or never"; 2 = "Sometimes"; 3 = "Much of the time"; 4 = "Just about all of the time"). At the time of measure development, the parent version of the COSS

Table 1 Demographic characteristics and COSS score means of ADHD sample across studies

Variable (percent of sample)	Full sample ($N = 619$)	Study 1 ($n = 325$)	Study 2 ($n = 237$)	Study 3 ($n = 57$)
Mean age (in years)	12.1 ± 1.0	12.3 ± 0.9	12.0 ± 1.1	11.9 ± 1.0
White/Caucasian	424 (69%)	251 (77%)	133 (56%)	40 (70%)
Black/African-American	127 (21%)	39 (12%)	72 (30%)	16 (28%)
Multiracial	54 (9%)	27 (8%)	26 (11%)	1 (2%)
Male	446 (72%)	231 (71%)	171 (72%)	44 (77%)
ADHD—combined presentation	272 (44%)	158 (49%)	88 (37%)	26 (44%)
Using medication for ADHD treatment	324 (52%)	153 (47%)	137 (58%)	34 (61%)
COSS Total Score	133.1 ± 12.8	133.4 ± 13.7	132.6 ± 12.1	134.2 ± 9.9
COSS Task Planning	15.9 ± 4.2	16.3 ± 4.2	15.9 ± 4.0	13.1 ± 4.3
COSS Organized Actions	15.8 ± 4.0	15.3 ± 3.5	15.8 ± 3.6	19.9 ± 6.7
COSS Memory/Materials	24.0 ± 5.7	24.9 ± 5.5	23.4 ± 5.6	21.0 ± 6.2

demonstrated high internal consistency ($\alpha s = 0.89$–0.91) and high test–retest reliability ($\alpha s = 0.94$–0.99) across all three scales. To evaluate convergent and divergent validity, COSS scores were compared to ratings on the Conners 3 (Conners, 2008), which is designed to evaluate ADHD symptoms and related areas of impairment. The COSSs were highly correlated with ADHD symptoms, learning problems, and executive dysfunction as measured by the Conners 3 (rs = 0.48–0.70). In contrast, COSS scores were not consistently correlated with conduct or peer relationship problems (rs = 0.04–0.36). Finally, the COSS exhibited good discriminant validity between students with ADHD, students with learning disorders, and typically developing students. The results of the discriminant validity tests found that students with ADHD generally exhibited more OTMP deficits than typically developing students or students with a learning disorder (Abikoff & Gallagher, 2009). Studies using the COSS as an outcome measure have found that it is sensitive to changes in behavior after children receive targeted intervention (e.g., Evans et al., 2016; Pfiffner et al., 2007).

The COSS produces a Total Score, which is a sum of all 58 items within the form. Within the 58 items, 26 of the items are part of one of three scales established by the CFA published in the technical manual: Task Planning (TP; 6 items), Organized Actions (OA; 10 items), and Memory/Materials Management (MM; 10 items). The Task Planning scale assesses a youth's ability to develop a strategy for completing assignments, chores, and other projects using items such as "My child seems to run out of time before school assignments are finished." The Organized Actions scale evaluates a child's execution of strategies or routines and their use of tools to complete goals. Example items in this scale include "My child puts his/her things in order" and "To help with written work, my child first makes a rough draft or an outline." The Memory/Materials Management scale assesses how well youth keep track of both physical belongings (e.g., books and clothes) as well as appointments and due dates using items such as "My child loses things at school" (Abikoff & Gallagher, 2009). As mentioned previously, the remaining items that contribute to the Total Score, but not one of the scale scores were retained for one of two reasons. First, they may contribute to the Inconsistency Index (10 item pairs) to evaluate whether a rater responds to similar items in the same way. Alternatively, they were retained on the form because they held high clinical relevance; 26 items appear to have been retained for this reason, although the exact items retained for this reason were not reported. It should also be noted that no justification was given for retaining any specific item that did not load on one of the three established factors or that was part of the Inconsistency Index (Abikoff & Gallagher, 2009).

The internal consistency values calculated from the current study data were lower than those reported in the COSS manual; the Task planning ($\alpha = 0.80$) and Memory/Materials Management ($\alpha = 0.83$) scales demonstrated high internal consistency, but the Organized Action scale ($\alpha = 0.64$) and Total Score ($\alpha = 0.74$) only demonstrated adequate internal consistency. However, alpha coefficients have been found to be variable across studies using the COSS, with values ranging from 0.64 to 0.90 (Evans et al., 2016; Pfiffner et al., 2007). The mean scores and standard deviations for these three scales and the Total Score in the full sample and across sites are available in Table 1.

During the exploratory factor analysis, the identified three-factor solution originally accounted for 62% of the variance in the data. Matching models were then created for the parent and child forms. Confirmatory factor analysis conducted on each version of the COSS found that the model adequately fit the data from the full COSS measure development sample. These judgments were made using the comparative fit index (CFI; 0.94–0.96), Tucker–Lewis index (TLI; 0.92–0.95), and root-mean-square error of approximation (RMSEA; 0.06–0.11) (Abikoff & Gallagher, 2009).

Also during scale development, a total of 11 item parcels were created consisting of either 2 or 3 items per parcel; 3 parcels were included in the TP factor, and 4 parcels each were included in the OA and MM factors. The developers noted that the item parcels were generally created due to high inter-correlations between individual items. The correlation coefficients between the items were not reported, but the rationale has some face validity. For example, one Task Planning parcel contains the items "Once my child gets ready to do schoolwork or projects, he/she has trouble knowing how to start" and "When my child has a big project to do, he/she doesn't know where to begin." The exact item parcels are available in the technical manual, and a brief overview of the parcels is available in Fig. 1 of the current study.

Other Measures

Behavior Rating Inventory of Executive Function—Parent Report (BRIEF; Gioia et al. 2000a)

The BRIEF—parent report—is an 86-item rating scale designed for the assessment of EF in children 5–18 years of age. Valid completion of all items results in the formulation of eight clinical scales, which in turn are grouped into two broad index sores and one total composite score. The eight clinical scales included in the BRIEF are Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. Each clinical scale is derived from at least 6 items. The Inhibit, Shift, and Emotional Control scales are combined to create a score referred to as the Behavior Regulation Index (BRI), while the remaining scales are used to create the Metacognition Index (MI). All eight clinical scales are used to calculate the General Executive Composite (GEC) score. The BRIEF is psychometrically valid with adequate internal consistency (αs = 0.80–0.98), test–retest reliability (0.76–0.88), and construct validity established through convergent and discriminant analyses (Gioia et al., 2000b; Mahone et al., 2002; McCandless & O'Laughlin, 2007).

Analytic Plan

All factor analyses were completed using Mplus version 7 (Muthén & Muthén, 1998–2012). The 26 items that made up the three COSS scales were used as indicators in each of the models. The amount of missing data in analyses was negligible, with only 2.8% of data missing from all item responses across participants. Given the discrete number of responses allowed on the COSS, all items were treated as ordered categorical indicators and the weighted least squares (WLSMV) estimator was used for all models. Evaluation of the relative fit of all tested models was conducted using both comparative and absolute fit indices. Specifically, the CFI, TLI, and RMSEA were examined using value threshold criteria recommended by Hu and Bentler (1999) for assessing model fit. Based on these recommendations, a model would be deemed to have acceptable fit if the CFI and TLI values approached or exceeded 0.95, and the RMSEA value approached or fell below 0.06.

Three total models were conducted: two models to evaluate the first study aim and one model to evaluate the second study aim. The first model tested was a replication of the CFA reported in the COSS technical manual. In this model, a three-factor solution was tested using the item parcels created in the original CFA. The three factors were allowed to correlate with one another (i.e., an oblique orientation). The second model tested used the same three-factor solution as the first model, with the exception that the indicators associated with each factor were not parceled. This model allowed for an examination of the effect of item parceling on the model fit index values for this factor structure. As these analyses were designed to be replications of previous work, no post hoc modifications were made to the models.

To complete the second study aim, a bi-factor model was estimated in which all indicators were loaded onto both their original domain-specific factor and a broader domain-general factor. Because the domain-general factor theoretically accounts for shared variance across all indicators, the variance to be accounted for in the domain-specific factors should be minimal (Reise, 2012); therefore, the model specified these factors to be orthogonal (i.e., uncorrelated with one another). The factor loadings for each item were also evaluated to determine how strongly individual items loaded onto the general or specific factors. Figure 2 presents a visual depiction of the proposed bi-factor model. Additionally, two measures of internal validity for a domain-general factor were calculated to examine whether one general score or several specific scores would provide more reliable information about OTMP skills. Procedures for calculating both measures are outlined by Rios and Wells (2014). First, omega hierarchical coefficients for the domain-general factor (ω_H) and each domain-specific factor (ω_s) were calculated to determine whether Total Scores and/or subscale scores provide unique and reliable information after controlling for the other factors. Second, the explained common variance (ECV) was calculated to provide a measure of the proportion of total explained variance that is accounted for by a single factor. For both omega hierarchical coefficients

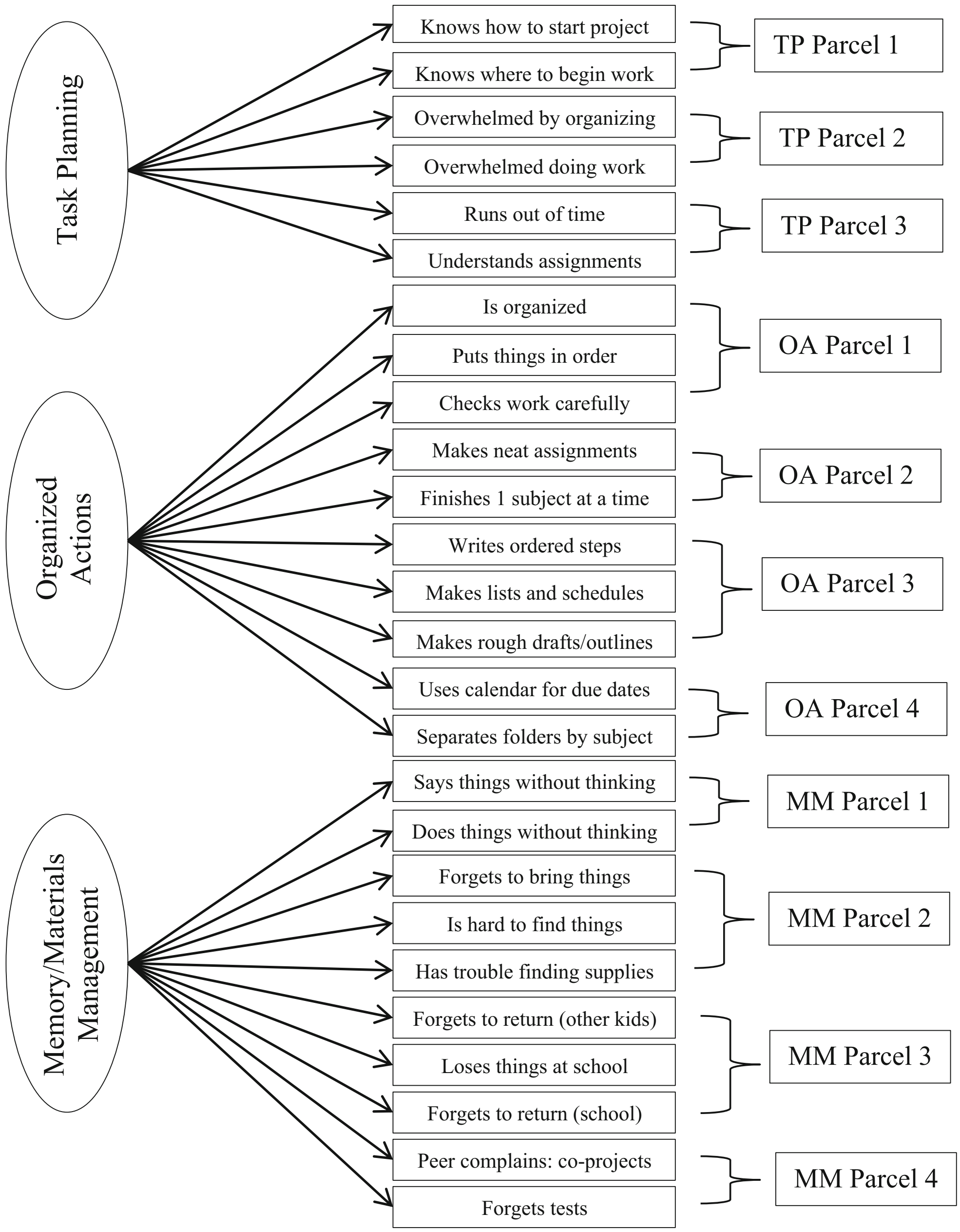

Fig. 1 Original 3-factor structure proposed by Abikoff and Gallagher (2009)

and ECVs, higher values indicate that a particular factor makes a stronger contribution to the overall variance in COSS ratings accounted for by the model (Rios & Wells, 2014).

The third study aim focused on examining correlations between the domain-general Total Score suggested by the bi-factor model and several scores generated by the BRIEF—parent report. These analyses used only a subsample of the full dataset ($n = 237$) in which parent ratings were collected for both the BRIEF and COSS. First, domain-general Total Scores were correlated with the GEC, BRI, and MI scores. Total Scores were also correlated with specific scale scores that should also measure OTMP skills specifically: Plan/Organize and Organization of Materials.

Results

Table 2 presents a comparison of the Chi-square goodness-of-fit values, degrees of freedom, and the fit statistics selected for model evaluation for each of the models tested

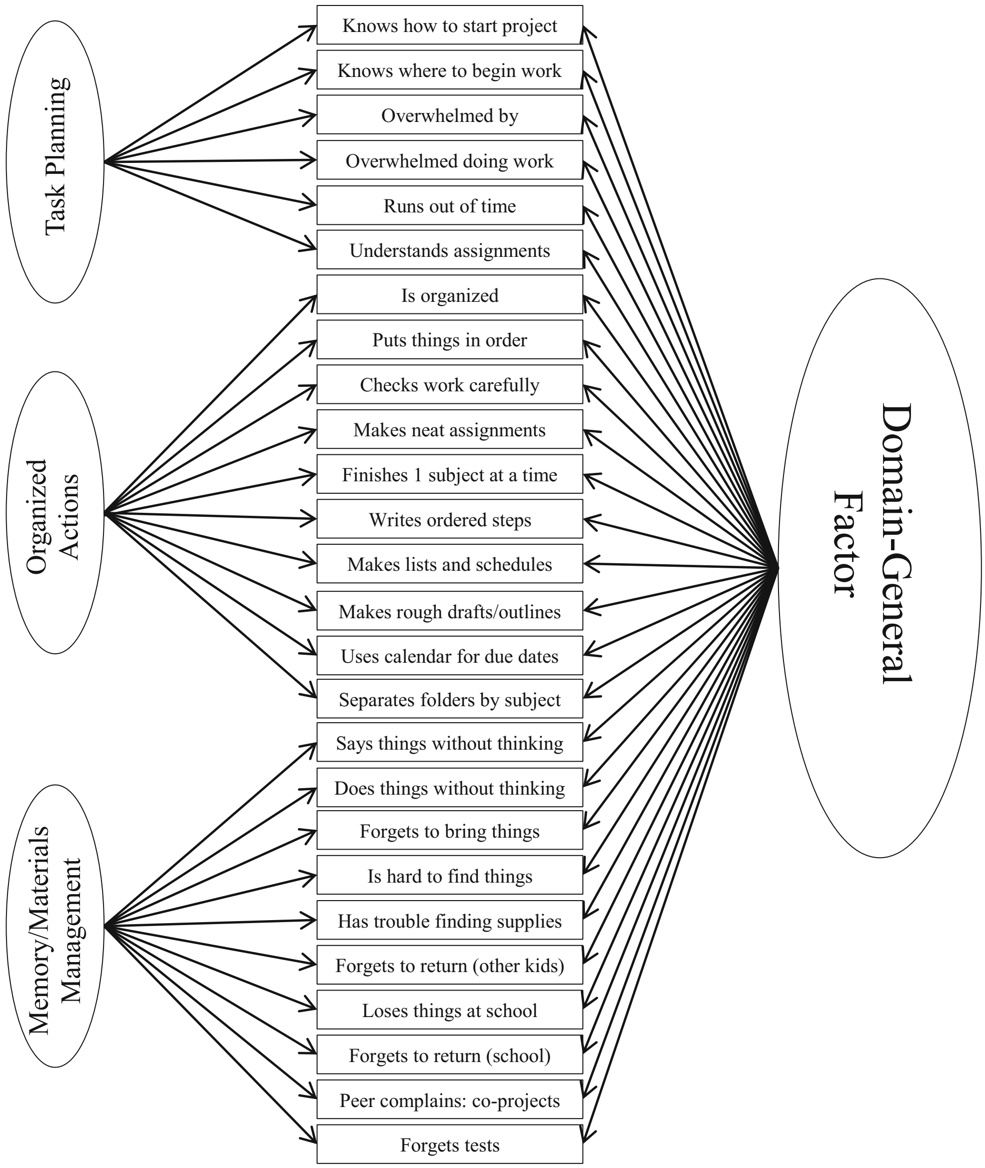

Fig. 2 Proposed bi-factor model

Table 2 Fit indices from confirmatory analyses of parent report of children's organizational skills scale

Model	χ^2	*df*	RMSEA	CFI	TLI	WRMR
Original model with parcels	666.78	277	0.050	0.948	0.938	1.197
Original model without parcels	1573.38	296	0.085	0.834	0.818	1.895
Bi-factor model without parcels	664.20	273	0.051	0.947	0.937	1.155

For CFI and TLI, higher values indicate better model fit. For RMSEA and WRMR, lower values indicate better model fit

RMSEA root-mean-square error of approximation, *CFI* comparative fit index, *TLI* Tucker–Lewis index, *WRMR* weighted root-mean-square residual

in this study. When the original factor structure presented in the technical manual was tested, the fit statistics indicated an acceptable model fit based on CFI (0.948), TLI (0.938), and RMSEA (0.050) values. Importantly, inter-factor correlations were high between all factors (*r* values = 0.41, 0.59, and 0.60), which may suggest that all items in the model may share a significant amount of variance with a common factor not included in the model.

When the second model—the original factor structure without any item parcels—was tested, the fit index values dropped below the established criteria (CFI = 0.834, TLI = 0.818, RMSEA = 0.085). This reduction in fit suggests that the item parceling procedure inflated the fit statistics of the original model. Again, the inter-factor correlations were high between all factors (r values = 0.47, 0.67, and 0.72). When the bi-factor model was tested, the fit statistics were acceptable and similar to those of the original model (CFI = 0.947, TLI = 0.937, RMSEA = 0.051), although no item parceling was used. Based on these fit indices and the lack of item parceling, the bi-factor model was determined to be the optimal solution from the alternative factor structures.

Table 3 presents the standardized factor loadings, expected common variance, and hierarchical omega coefficients for the final bi-factor solution. It should be noted that because of the bi-factor model specifications—specifically that the domain-specific factors are uncorrelated with one another—cross-loadings are only reported for the domain-general factor and the specific factor on which each item loaded. There was a considerable range in item loadings across the domain-general factor (0.20–0.80), although 14 of the 26 items loaded at least 0.40 onto the general factor. Similarly, the loadings for the Task Planning (0.38–0.67), Organized Actions (0.13–0.73), and Memory/Materials Management (0.01–0.80) domain-specific scales exhibited a wide range of values. Interestingly, four items from the Memory/Materials management subscale did not significantly load onto the domain-specific factor after controlling for their loadings on the domain-general factor (forgets about tests, forgets to bring things to school, hard for student to find papers, trouble finding supplies when needed). All other items significantly loaded

Table 3 Standardized loadings and internal validity values for bi-factor model

Item	General factor	Task Planning	Organized Actions	Memory/Material Management
Trouble starting schoolwork	0.46	0.65		
Runs out of time	0.48	0.38		
Doesn't know how to begin big projects	0.43	0.67		
Overwhelmed trying to stay organized	0.44	0.53		
Doesn't understand directions	0.55	0.40		
Overwhelmed trying to get work done on time	0.37	0.57		
Is organized	0.44		0.49	
Puts things in order	0.34		0.70	
Makes lists/reminders	0.20		0.73	
Checks work carefully	0.34		0.47	
Assignments are neat	0.33		0.36	
Finishes one subject before starting another	0.38		0.13	
Writes down steps	0.27		0.50	
Uses a calendar	0.29		0.50	
Makes rough drafts/outlines	0.25		0.30	
Uses separate folders for different subjects	0.37		0.18	
Forgets about tests	0.68			0.01
Forgets to bring things to school	0.80			0.01
Hard to find papers	0.87			0.03
Forgets to return things to other kids	0.51			0.34
Does things without thinking	0.36			0.80
Loses things at school	0.64			0.14
Says things without thinking	0.22			0.76
Forgets to return things to school	0.49			0.18
Others complain child is not organized	0.30			0.13
Trouble finding supplies when needed	0.78			0.03
Hierarchical omega coefficient (ω)	0.75	0.07	0.12	0.04
Explained common variance	0.53	0.15	0.19	0.12

Cross-loadings for items across domain-specific factors are not reported due to the bi-factor model assumption that the domain-specific factors are orthogonal onto both the domain-general factor and their respective domain-specific factors. The omega reliability coefficients indicated that the domain-general factor exhibited acceptable reliability ($\omega_H = 0.75$), whereas the coefficients for the Task Planning ($\omega_s = 0.07$), Organized Actions ($\omega_s = 0.12$), and Memory/Materials Management ($\omega_s = 0.04$) scales indicated that none of the domain-specific factors exhibited acceptable reliability after the influence of the domain-general factor was accounted for. Similarly, the explained common variance value for the domain-general factor was moderately high (ECV = 0.53) and low for each of the domain-specific factors (ECVs = 0.12–0.19). Together, these internal validity values support the presence of a strong domain-general factor within the COSS factor structure.

In order to test the convergent validity of the domain-general factor, a Total Score (e.g., item sums) of the 26 items included in the model was calculated for a subset of participants within the full sample for whom parent BRIEF ratings were also gathered. Bivariate correlations were calculated between the 26-item COSS score and the standardized T scores for all BRIEF scales, and these correlations are presented in Table 4. The domain-general Total Score was moderately correlated with the GEC ($r = 0.55$) from the parent-rated BRIEF in the subsample of participants with data from both rating scales ($n = 237$). They were also moderately correlated with both BRI scores ($r = 0.42$) and MI scores ($r = 0.55$). Correlations between the domain-general Total Score and all eight BRIEF clinical scales were also statistically significant ($ps < 0.001$). The clinical scale with the strongest correlation—and the only one with a correlation above 0.50—was the Plan/Organize scale ($r = 0.52$). Interestingly, the Organization of Materials scale exhibited the second lowest correlation of the remaining clinical scales ($r = 0.32$).

Table 4 Bivariate correlations between domain-general factor Total Score and BRIEF scale scores

BRIEF scale	Correlation with COSS
General Executive Composite	0.55
Behavior Regulation Index	0.39
Metacognition Index	0.55
Inhibit	0.31
Shift	0.34
Emotion Control	0.33
Initiate	0.41
Working Memory	0.47
Plan/Organize	0.52
Organization of materials	0.32
Monitor	0.43

Discussion

The primary goal of this study was to evaluate the factor structure of parent ratings on the COSS in a sample of adolescents with ADHD ($N = 619$). First, the original factor structure presented in the COSS technical manual was replicated when item parceling was allowed. However, model fit became unacceptable when the items were not parceled. Second, an alternative, theoretically supported bi-factor model was tested and it exhibited acceptable fit without item parceling. Third, the construct validity of the domain-general factor was explored using a subsample of the full replication sample. The domain-general factor of the bi-factor model exhibited a significant, moderate association with the General Executive Composite of the BRIEF and, as hypothesized, had a stronger correlation with the Metacognition Index in comparison with the Behavior Regulation Index.

The replication of the original factor structure in a large ADHD sample is important given that students with ADHD have considerable difficulties with OTMP skills and that the three-factor structure of the COSS has been used as an outcome in multiple large randomized controlled trials. It is important to note, however, that when item parcels were not allowed to be part of the model, the fit indices significantly declined. Researchers opposed to item parceling in CFAs have argued that the technique should not be used because they may indicate good model fit even though the model is misspecified (Bandalos, 2002; Marsh et al., 2013), and it appears that parceling may have had that effect on the original factor structure. Proponents of item parceling argue that if a sound theoretical reason for combining items exists and if parcels are determined before conducting the CFA, then item parceling is an appropriate technique (Little et al., 2013; Matsunaga, 2008). We could not determine whether this rationale applied to the analyses reported by the developers.

Regardless of item parceling, the bi-factor model conducted in this study presents an interesting, theory-driven alternative to the original model. Specifically, the bi-factor model fit was nearly identical to the original model even though item parceling was not used. Item parceling relies upon item relationships at two levels: between the items and the latent factor and between items within a parcel. However, the within-parcel relationships are frequently un-modeled making it difficult to disentangle the contributions of individual items. In contrast, the bi-factor model explicitly models the item relationships between both the domain-general and domain-specific factors, allowing for the evaluation of individual item contributions to all factors. For this reason, the bi-factor model allows for a more in-depth examination of the relationship of single items to COSS factor scores or to other related measures.

In the present study, the general factor demonstrated acceptable reliability ($\omega_H = 0.75$), especially in comparison with the domain-specific factors, and accounted for a moderately high portion of the common variance in COSS ratings among all factors (ECV = 0.53). This empirical support is paired with strong theoretical support for the presence of a domain-general factor within the COSS, as OTMP skills are rarely used in isolation. For example, successfully completing a task such as a homework assignment requires a student to remember the assignment, have all necessary materials to complete the assignment, and execute a plan to complete the assignment (Langberg et al., 2016). OTMP skills are also considered part of the set of broader cognitive abilities known as executive functions, and these functions often act in unison to complete a task or achieve a goal (Pennington & Ozonoff, 1996). Some of the items within the COSS appear to even encourage a broader evaluation of a child's skills (e.g., "My child is organized"). Therefore, incorporating a factor that can account for the broad connection between these items would increase the value of the data collected by the COSS.

Besides the construct validity demonstrated during the factor analysis, the convergent validity analyses suggest that the COSS domain-general factor may be better conceptualized as an indicator of cool EF abilities broadly instead of OTMP skills only. As expected, the domain-general Total Score was moderately correlated with the General Executive Composite (Table 4). Also consistent with hypotheses, the correlation magnitude was stronger between the COSS Total Score and the BRIEF Metacognition Index scores than Behavior Regulation Index scores, which indicates that the domain-general factor may be more representative of the "cool" end of the EF spectrum. The clinical scale scores provided more mixed evidence of convergent validity. The Plan/Organize subscale demonstrated a moderate correlation with the domain-general score, but the Organization of Materials only exhibited a modest correlation. More broadly, none of the clinical scale correlations with the COSS appeared substantially stronger from the rest of the scores, suggesting that the domain-general Total Score is associated with a wide range of EF abilities. However, these analyses were not intended to serve as a comprehensive validity assessment and additional work is necessary to demonstrate the convergent and discriminant validity of the domain-general factor. Future studies should also evaluate whether the domain-general Total Score is more strongly related to functional outcomes of interest (e.g., academic functioning) or whether it is more sensitive to treatment change than the originally suggested 58-item Total Score.

Although the bi-factor model makes a subtle change to the previously proposed structure of the COSS, the implications for how researchers and clinicians should use the tool in the future are significant. Perhaps one of the more significant recommendations regards the interpretability of the specific scale scores. Based upon the ECV values, it was clear that the general factor accounted for a substantially larger portion of the total variance in COSS ratings than the domain-specific scales. The reliability of all specific scales, as indicated by the omega hierarchical values, was also low when the effect of the general factor was accounted for. Additionally, the item loadings on the specific factors, especially in comparison with item loadings on the general factor, appear to call into question the interpretability of the specific factors. Although the items of the Task Planning factor all loaded at least modestly (0.38–0.67) onto the specific factors, both the Organized Actions and the Memory/Materials Management scales had multiple items with loadings below 0.20, suggesting that those items contribute little to the specific factor. Together, these findings suggest that a single Total Score from all items used in the bi-factor model would be a more reliable indicator of OTMP skills than scores summed only from items within a specific factor.

A second implication of the findings is that a Total Score calculated from fewer items may still provide a reliable measure of OTMP skills. The COSS technical manual encourages the calculation of a Total Score from 58 items, excluding only the impairment items (Abikoff & Gallagher, 2009). However, both the original three-factor structure and the bi-factor model only use 26 of those items. Six of the items within the COSS that are not included are part of the 20 items used to calculate the Inconsistency Index. The purpose of the remaining 26 items is not clear. In the technical manual, the authors refer to retaining a set of seven items that loaded onto multiple factors "…because each item pertained to behaviors or outcomes believed to be clinically important" (p. 48, Abikoff & Gallagher, 2009). The exact seven items that were affected by this decision were not specified, and it is unclear whether the remaining 13 items that are not included in the measurement models were also retained for their clinical significance or whether they were retained for other reasons. Further, it is possible that these items may have a negative influence on the reliability and validity of the Total Score calculated from the COSS. Because these items are not included in any factor analyses to date and their loadings were not reported during original scale development, it is unclear how closely responses on these items align with other items already included in the factor structure. The COSS would likely benefit from the creation of a brief version of the measure, which would save time during research protocols or psychological assessments and reduce the burden placed on parents completing the rating scale. The bi-factor model provides a useful starting point for considering the development of such a measure, as it relies upon only half of the items within the current version of the COSS. Alternatively, the additional 30 items may capture unique and important variance in functioning and may even have the potential to be captured by currently unidentified domain-specific factors. However, significantly more research would be necessary before they could be reasonably incorporated into either the factor structure proposed by Abikoff and Gallagher (2009) or into the proposed bi-factor model.

Although the current study offers important insights into the validity of the COSS, several limitations should be noted. First, the original COSS development sample captured a broader age range than the current sample. Specifically, the original COSS sample included some younger students in elementary grades. Similarly, the current study sample consisted entirely of youth formally diagnosed with ADHD, whereas the original COSS validation study included both general and clinical samples. Therefore, it is possible that the factor structure of the COSS varies across development or across clinical presentations and that an alternative factor structure would more accurately represent ratings for younger or typically developing students. In addition, it is possible that additional factors such as gender or ADHD medication status could impact the factor structure. Although the gender distribution and ADHD medication rates in our sample are consistent with other large ADHD samples followed into early adolescence (e.g., Jensen et al., 2007; Visser et al., 2013), it would still be valuable to evaluate potential differences in ratings within these subgroups. Unfortunately, sample size of the current study was not sufficiently large to examine factor invariance (i.e., applicability of the factor structure) across these groups. In summary, additional replication and invariance testing of the identified factor structure is needed to explore the potential impact of age, sample type, gender, and medication status.

A second limitation is that the COSS ratings were limited to parent report in this study. Although a teacher version of the COSS is available, there are significant methodological hurdles that make comparison between parent and teacher ratings difficult in the middle school context. Indeed, parent/teacher agreement on ratings of ADHD symptoms and impairment is often low in secondary school settings (Evans et al., 2005; Narad et al., 2015), and each student has multiple teachers who may disagree about student functioning (Fabiano, Chafouleas, Weist, Sumi, & Humphrey, 2014). Nonetheless, because both parent and teacher reports on the COSS are used in ADHD intervention studies, the factor structure of the COSS teacher version should be evaluated. Finally, although it was determined that the item parcels of the Abikoff and Gallagher (2009) model inflated fit statistics, the specific mechanisms driving the inflation were not explored in the current study. Future research could investigate the possible underlying mechanisms, which would lead to the creation of more accurate, parsimonious factor structures.

In summary, this study presents an evaluation of the factor structure of the COSS parent version. The study found that the original factor structure could successfully be replicated, although item parceling appears to inflate the model's fit. An alternative bi-factor model demonstrated acceptable fit to the data, and as such, the presence of a domain-general factor within OTMP skills warrants future research attention across raters and in relation to outcomes.

Acknowledgements This research was supported by Grant Number R01 MH082865 awarded to Steven W. Evans by the National Institute of Mental Health, and Grant Number R305A130011 awarded to Joshua M. Langberg by the Institute of Education Sciences.

Compliance with Ethical Standards

Conflict of interest Stephen J. Molitor, Joshua M. Langberg, Steven W. Evans, Melissa R. Dvorsky, Elizaveta Bourchtein, Laura D. Eddy, Zoe R. Smith, and Lauren E. Oddo declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

Abikoff, H., & Gallagher, R. (2009). *COSS: Children's organizational skills scales*. North Tonawanda, NY: Multi-Health Systems.

Abikoff, H., Nissley-Tsiopinis, J., Gallagher, R., Zambenedetti, M., Seyffert, M., Boorady, R., et al. (2009). Effects of MPH-OROS on the organizational, time management, and planning behaviors of children with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*(2), 166–175.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Arlington Country: American Psychiatric Publications.

Anesko, K. M., Schoiock, G., Ramirez, R., & Levine, F. M. (1987). The homework problem checklist: Assessing children's homework difficulties. *Behavioral Assessment, 9*, 179–185.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*(1), 78–102.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Hillsdale, NJ: Erlbaum.

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2007). Long-term school outcomes for children with attention-deficit/hyperactivity disorder: A population-based perspective. *Journal of Developmental and Behavioral Pediatrics, 28*(4), 265–273.

Barkley, R. A. (1997). Attention-deficit/hyperactivity disorder, self-regulation, and time: Toward a more comprehensive theory. *Journal of Developmental and Behavioral Pediatrics, 18*(4), 271–279.

Booster, G. D., DuPaul, G. J., Eiraldi, R., & Power, T. J. (2010). Functional impairments in children with ADHD: Unique effects of age and comorbid status. *Journal of Attention Disorders*. doi:10.1177/1087054710383239.

Conners, C. K. (2008). *Conners* (3rd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

DuPaul, G. J., & Langberg, J. L. (2014). Educational impairments in children with ADHD. In R. A. Barkley (Ed.), *Attention-deficit hyperactivity disorder (4th Ed) A handbook for diagnosis and treatment*. New York: Guilford Press.

Evans, S. W., Allen, J., Moore, S., & Strauss, V. (2005). Measuring symptoms and functioning of youth with ADHD in middle schools. *Journal of Abnormal Child Psychology, 33*(6), 695–706.

Evans, S. W., Langberg, J. M., Schultz, B. K., Vaughn, A., Altaye, M., Marshall, S. A., et al. (2016). Evaluation of a school-based treatment program for young adolescents with ADHD. *Journal of Consulting and Clinical Psychology, 84*(1), 15.

Evans, S. W., Schultz, B. K., White, L. C., Brady, C., Sibley, M. H., & Van Eck, K. (2009). A school-based organization intervention for young adolescents with attention-deficit/hyperactivity disorder. *School Mental Health, 1,* 78–88.

Fabiano, G. A., Chafouleas, S. M., Weist, M. D., Sumi, W. C., & Humphrey, N. (2014). Methodology considerations in school mental health research. *School Mental Health, 6*(2), 68–83.

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: Meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities, 40*(1), 49–65.

Fristad, M. A., Teare, M., Weller, E. B., Weller, R. A., & Salmon, P. (1998). Study III: Development and concurrent validity of the children's interview for psychiatric syndromes- parent version (P-ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8*(4), 221–226.

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

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000b). Test review: Behavior rating inventory of executive function. *Child Neuropsychology, 6,* 235–238.

Gureasko-Moore, S., DuPaul, G. J., & White, G. P. (2007). Self-management of classroom preparedness and homework: Effects on school functioning of adolescents with attention deficit hyperactivity disorder. *School Psychology Review, 36*(4), 647.

Harrier, L. K., & DeOrnellas, K. (2005). Performance of children diagnosed with ADHD on selected planning and reconstitution tests. *Applied Neuropsychology, 12*(2), 106–119.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416.

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.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health, 7*(3), 131–142.

Jensen, P. S., Arnold, L. E., Swanson, J. M., Vitiello, B., Abikoff, H. B., Greenhill, L. L., et al. (2007). 3-Year follow-up of the NIMH MTA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(8), 989–1002.

Langberg, J. M., Dvorsky, M. R., & Evans, S. W. (2013). What specific facets of executive function are associated with academic functioning in youth with attention-deficit/hyperactivity disorder? *Journal of Abnormal Child Psychology, 41*(7), 1145–1159.

Langberg, J. M., Dvorsky, M. R., Molitor, S. J., Bourchtein, E., Eddy, L. D., Smith, Z., et al. (2016). Longitudinal evaluation of the importance of homework assignment completion for the academic performance of middle school students with ADHD. *Journal of School Psychology, 55,* 27–38.

Langberg, J. M., Epstein, J. N., Becker, S. P., Girio-Herrera, E., & Vaughn, A. J. (2012). Evaluation of the Homework, Organization, and Planning Skills (HOPS) intervention for middle school students with ADHD as implemented by school mental health providers. *School Psychology Review, 41*(3), 342.

Langberg, J. M., Epstein, J. N., Girio-Herrera, E., Becker, S. P., Vaughn, A., & Altaye, M. (2011). Materials organization, planning, and homework completion in middle-school students with ADHD: Impact on academic performance. *School Mental Health, 3*(2), 93–101.

Leffler, J. M., Riebel, J., & Hughes, H. M. (2014). A review of child and adolescent diagnostic interviews for clinical practitioners. *Assessment*. doi:10.1177/1073191114561253.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*(3), 285.

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology, 32*(6), 643–654.

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., et al. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology, 17*(7), 643–662.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J., & 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.

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures, 2*(4), 260–293.

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,* 381–389.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (*7th Ed.*). Los Angeles, CA: Muthén & Muthén.

Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., et al. (2015). Parent–teacher agreement on ADHD symptoms across development. *Psychological Assessment, 27*(1), 239.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*(1), 51–87.

Pfiffner, L. J., Mikami, A. Y., Huang-Pollock, C., Easterlin, B., Zalecki, C., & McBurnett, K. (2007). A randomized, controlled trial of integrated home-school behavioral treatment for ADHD, predominantly inattentive type. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(8), 1041–1050.

Pfiffner, L. J., Villodas, M., Kaiser, N., Rooney, M., & McBurnett, K. (2013). Educational outcomes of a collaborative school–home behavioral intervention for ADHD. *School Psychology Quarterly, 28*(1), 25.

Power, T. J., Dombrowski, S. C., Watkins, M. W., Mautone, J. A., & Eagle, J. W. (2007). Assessing children's homework performance: Development of multi-dimensional, multi-informant rating scales. *Journal of School Psychology, 45*(3), 333–348.

Power, T. J., Watkins, M., Mautone, J. A., Walcott, C. M., Coutts, M. J., & Sheridan, S. M. (2015). Examining the validity of the Homework Performance Questionnaire: Multi-informant assessment in elementary and middle school. *School Psychology Quarterly, 30*(2), 260–275.

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology, 108*(3), 621–637.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696.

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema, 26,* 108–116.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 298*(1089), 199–209.

Shimoni, M. A., Engel-Yeger, B., & Tirosh, E. (2012). Executive dysfunctions among boys with Attention Deficit Hyperactivity Disorder (ADHD): Performance-based test and parents report. *Research in Developmental Disabilities, 33*(3), 858–865.

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.

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, M., Kogan, R., Ghandour, R., et al. (2013). Trends in the prevalence of parent-reported ADHD diagnosis and medication treatment by a health care professional: United States, 2003–2011. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(1), 34–46.

Wechsler, D.(2003). *Wechsler intelligence scale for children-WISC-IV*. Psychological Corporation.

Weller, E. B., Weller, R. A., Fristad, M. A., Rooney, M. T., & Schecter, J. (2000). Children's interview for psychiatric syndromes (ChIPS). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(1), 76–84.

Weyandt, L., DuPaul, G. J., Verdi, G., Rossi, J. S., Swentosky, A. J., Vilardo, B. S., et al. (2013). The performance of college students with and without ADHD: Neuropsychological, academic, and psychological functioning. *Journal of Psychological Behavioral Assessment*. doi:10.1007/s10862-013-9351-8.

Wolraich, M. L., Wibbelsman, C. J., Brown, T. E., Evans, S. W., Gotlieb, E. M., Knight, J. R., et al. (2005). Attention-deficit/hyperactivity disorder among adolescents: A review of the diagnosis, treatment, and clinical implications. *Pediatrics, 115*(6), 1734–1746.

Zoromski, A. K., Owens, J. S., Evans, S. W., & Brady, C. E. (2015). Identifying ADHD symptoms most associated with impairment in early childhood, middle childhood, and adolescence using teacher report. *Journal of Abnormal Child Psychology, 43,* 1243–1255.