

Construct Validity of ADHD/ODD Rating Scales: Recommendations for the Evaluation of Forthcoming DSM-V ADHD/ODD Scales

G. Leonard Burns · James A. Walsh · Mateu Servera · Urbano Lorenzo-Seva · Esther Cardo · Antoni Rodríguez-Fornells

Published online: 8 July 2012
© Springer Science+Business Media, LLC 2012

Abstract Exploratory structural equation modeling (SEM) was applied to a multiple indicator (26 individual symptom ratings) by multitrait (ADHD-IN, ADHD-HI and ODD factors) by multiple source (mothers, fathers and teachers) model to test the invariance, convergent and discriminant validity of the Child and Adolescent Disruptive Behavior Inventory with 872 Thai adolescents and the ADHD Rating Scale-IV and ODD scale of the Disruptive Behavior Inventory with 1,749 Spanish children. Most of the individual ADHD/ODD

symptoms showed convergent and discriminant validity with the loadings and thresholds being invariant over mothers, fathers and teachers in both samples (the three latent factor means were higher for parents than teachers). The ADHD-IN, ADHD-HI and ODD latent factors demonstrated convergent and discriminant validity between mothers and fathers within the two samples. Convergent and discriminant validity between parents and teachers for the three factors was either absent (Thai sample) or only partial (Spanish sample). The application of exploratory SEM to a multiple indicator by multitrait by multisource model should prove useful for the evaluation of the construct validity of the forthcoming DSM-V ADHD/ODD rating scales.

G. L. Burns (✉)
Department of Psychology, Washington State University,
Pullman, WA 99164-4820, USA
e-mail: glburns@wsu.edu

J. A. Walsh
University of Montana,
Missoula, MT, USA

M. Servera
Department of Psychology & Research Institute on Health
Sciences, University of the Balearic Islands,
Palma, Spain

U. Lorenzo-Seva
CRAMC (Research Center for Behavior Assessment),
Department of Psychology, Universitat Rovira i Virgili,
Tarragona, Spain

E. Cardo
Son Llatzer Hospital & Research Institute on Health Sciences,
University of Balearic Islands,
Palma, Spain

A. Rodríguez-Fornells
Institutio Catalana de Recerca i Estudis Avançats,
Barcelona, Spain

A. Rodríguez-Fornells
Department of Ciències Fisiològiques II, University of Barcelona,
Barcelona, Spain

Keywords ADHD · Oppositional defiant disorder · Multitrait by multimethod matrix · Construct validity · ADHD/ODD rating scales, Confirmatory factor analysis · Exploratory structural equation modeling

Attention-deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) rating scales play important roles in the advancement of our understanding of ADHD and ODD. The ratings scales are used to evaluate the structural organization of the ADHD/ODD symptoms (e.g., Burns et al. 2001c). The scales are also used to determine if the ADHD-inattention (IN), ADHD-hyperactivity/impulsivity (HI) and ODD dimensions have unique external correlates (e.g., biological markers, risk factors, associated features and outcomes, Barkley 2011). Finally, the ADHD/ODD rating scales have important roles in the diagnostic process as well as in the evaluation of the effectiveness of treatments. Given the important roles of the scales in research on ADHD/ODD, it is critical that the construct validity of the scales be evaluated in a thorough manner (Burns and Haynes 2006).

Confirmatory Factor Analysis and the Construct Validity of ADHD/ODD Rating Scales

With the publication of the ADHD and ODD symptoms in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM, American Psychiatric Association 1980), confirmatory factor analysis (CFA) became a common procedure to use to evaluate the construct validity of the scales (see Table 1 in Bauermeister et al. 2010 for a list of the CFA studies). There were two reasons for this occurrence. First, an explicit list of ADHD/ODD symptoms allowed researchers to turn the symptoms into items on rating scales. Second, because the DSM implied a specific measurement model, CFA provided a useful procedure to test the validity of the DSM model as well as alternative models (e.g., Moura and Burns 2010; Toplak et al. 2009, 2012).

The analysis of the structure of ADHD/ODD scales with CFA requires the cross-loadings to be restricted to zero. If one or more of the symptoms have significant cross-loadings on a secondary factor, then the use of CFA results in two problems. The first problem is the failure to identify symptoms with weak discriminant validity. The second problem is the possibility of biased results. As noted by Asparouhov and Muthén (2009), “when non-zero cross-loadings are specified as zero, the correlation between factor indicators representing different factors is forced to go through their main factors only, usually leading to overestimated factor correlations and subsequent distorted structural relations (p. 398).” In other words, the more the cross-loadings depart from zero, the more the correlations among the ADHD-IN, ADHD-HI and ODD factors will be inflated to account for non-zero cross-loadings restricted to zero, thus yielding biased loadings and factor correlations.

Table 1 Model fit indices for the ADHD-IN, ADHD-HI, and ODD three-factor model

Group	χ^2	CFI	TLI	RMSEA (90 % CI)
Thai mothers ($n=872$)	514	0.981	0.975	0.035 (0.030–0.039)
Thai fathers ($n=872$)	574	0.980	0.974	0.039 (0.034–0.043)
Thai teachers ($n=872$)	456	0.980	0.974	0.031 (0.026–0.035)
Spanish mothers ($n=1423$)	851	0.972	0.963	0.041 (0.038–0.044)
Spanish fathers ($n=1380$)	796	0.975	0.968	0.040 (0.037–0.043)
Spanish teachers ($n=1749$)	1217	0.983	0.977	0.047 (0.044–0.050)

ADHD-IN attention-deficit/hyperactivity disorder—inattention, *ADHD-HI* attention-deficit/hyperactivity disorder—hyperactivity/impulsivity, *ODD* oppositional defiant disorder, *CFI* comparative fit index, *TLI* Tucker-Lewis index, *RMSEA* root mean-square error of approximation. *CI* confidence interval. *N* refers to the number of children or adolescents rated by each source. Degrees of freedom equal 250 for each analysis

Asparouhov and Muthén (2009) recommend the use of exploratory structural equation modeling (SEM) when it is not appropriate to restrict the cross-loadings to zero (e.g., with new measurement instruments when the discriminant validity of the items is not known). While this procedure a priori specifies an ADHD-IN, ADHD-HI and ODD three-factor model similar to CFA, the cross loadings are freely estimated (i.e., each symptom has a primary and two secondary loadings). The exploratory SEM procedure offers three advantages relative to CFA for the evaluation of a new ADHD/ODD rating scale—(1) easy identification of ADHD-IN, ADHD-HI and ODD symptoms with weak discriminant validity (i.e., strong loadings on a secondary factor); (2) a more accurate representation of a symptom’s relationship with its primary factor; and (3) more accurate correlations among the ADHD-IN, ADHD-HI and ODD factors. Exploratory SEM with a single source thus allows a more accurate evaluation of new ADHD/ODD rating scales than CFA. Exploratory SEM with *multiple sources*, however, provides an even more sophisticated evaluation of the construct validity of the ADHD/ODD scales. The next section outlines the benefits of the merger of exploratory SEM with multiple sources.

Construct Validity of ADHD/ODD Rating Scales Between Multiple Sources

The application of exploratory SEM to a multiple indicator (individual symptoms) by multitrait (ADHD-IN, ADHD-HI and ODD factors) by multisource (mothers, fathers and teachers sources) model allows answers to six questions relevant to the construct validity of ADHD/ODD rating scales. This procedure allows one to determine the (1) convergent and discriminant validity of the individual ADHD/ODD symptoms for each source (i.e., *Do the individual symptoms have substantial loadings on their primary factor with the loadings on the primary factor being larger than the loadings on the two secondary factors for mothers, fathers and teachers?*), (2) invariance of like-symptom loadings and thresholds between sources (i.e., *Does the measurement model for each factor remain invariant between mothers, fathers, and teachers?*), (3) invariance of like-factor means between sources (i.e., *Do mothers, fathers and teachers perceive equal levels of the ADHD-IN, ADHD-HI and ODD factors in the sample of children?*), (4) convergent and discriminant validity of the factors between sources (i.e., *Are same factor-different source correlations substantial as well as significantly larger than the different factor-different source correlations?*), (5) discriminant validity of the factors within sources (i.e., *Are the different factor-same source correlations low enough to indicate meaningfully different symptom dimensions within each source?*) and (6) the magnitude of the source effects (i.e., *How much larger are the different factor-same*

source correlations than the different factor-different source correlations?). Answers to these six questions yield a great deal of information on the strength of an ADHD/ODD scale's construct validity.

Purpose of the Study

Our primary purpose was to demonstrate the application of exploratory SEM to a multiple indicator (26 individual symptoms) by multitrait (ADHD-IN, ADHD-HI and ODD factors) by multisource (mothers, fathers and teachers) model to show the merits of this procedure for the evaluation of the construct validity of the forthcoming DSM-V ADHD/ODD rating scales. Our secondary purpose was to use this procedure to evaluate the construct validity of two different DSM-IV ADHD/ODD rating scales (i.e., the Child and Adolescent Disruptive Behavior Inventory with Thai adolescents and the ADHD Rating Scale-IV combined with the ODD section of the Disruptive Behavior Disorders Rating Scale with Spanish children). To the best of our knowledge, this procedure has never been used to evaluate the construct validity of an ADHD/ODD scale between mothers, fathers and teachers' ratings.

Method

Participants and Procedures

Thai sample The participants were the mothers, fathers, and teachers of 872 Thai adolescents (7th to 12th grade) from the Demonstration School in the city of Mahasarakham (population approximately 117,500) in northeastern Thailand. With the approval of the school and Washington State University's IRB, the mothers, fathers, and teachers of the adolescents were invited to participate in the study. A total of 29 of 29 teachers volunteered to participate in the study with mothers' (95 % mothers and 5 % other relatives) and fathers' (91 % fathers and 9 % others relatives) ratings also being obtained on the 872 adolescents. The average age of the adolescents was 14.99 years ($SD=1.77$) with 61 % of the sample being female. The average educational level of the mothers and fathers was 14.66 ($SD=3.13$) and 14.53 ($SD=3.53$) grades, respectively. The average number of adolescents rated by each teacher was 30.07 ($SD=8.23$). No adolescents were excluded from the study. Information was not available on the number of adolescents receiving special services.

Spanish Sample The participants were the mothers, fathers, and teachers of 1,749 Spanish children (1st to 4th grade) from 16 randomly selected elementary schools from a total

215 schools on the island of Majorca in the Balearic Islands. The inclusion criteria for the children of potential participants were that the children could not have a school diagnosis of mental retardation, developmental coordination disorders, pervasive developmental disorders or severe emotional disturbance. This procedure resulted in 1,785 children as potential participants in the 80 randomly selected classes. With the approval of the schools and the IRB of the University of the Balearic Islands, the 1,785 families were invited to participate. A total of 36 families declined to participate. Teacher ratings were obtained on 1,749 children (1,785 – 36=1,749) with mothers' and fathers' ratings obtained on 1,422 and 1,380 of the 1,749 children, respectively. The average age of children was 8.31 years ($SD=1.21$) with 48 % of the sample being female. A total of 80 teachers participated in the study with each teacher rating an average of 21.87 children ($SD=11.88$).

Measures

Thai Sample—Child and Adolescent Disruptive Behavior Inventory (CADBI) Mothers, fathers and teachers rated the occurrence of the nine ADHD-IN, nine ADHD-HI, and eight ODD symptoms on an 8-point frequency-of-occurrence scale for the past month (i.e., 1 = *never in the past month*; 2 = *one to two times in the past month*; 3 = *three to four times in the past month*; 4 = *two to six times per week (or two to four times per week for teachers)*; 5 = *one time per day*; 6 = *two to five times per day*; 7 = *six to nine times per day*; 8 = *10 or more times per day*) (Burns et al. 2001a, b). Parents were instructed to make their ratings on the basis of the adolescent's behavior in the home and community and not to consider the adolescent's behavior toward teachers and peers at school. Mothers and fathers were also instructed to make their rating independently. Teachers were told to base their ratings on only the adolescents' behavior at school. The teachers had been interacting with the adolescents for almost the entire school year. The parents and teachers' ratings occurred at the same time.

Psychometric Properties Several studies support the construct validity of the parent and teacher versions of the CADBI. For example, the parent scale has demonstrated invariance of like-item loadings, intercepts and residuals as well as invariance of like-factor variances, covariances and factor means between mothers and fathers rating of the same child within samples of American, Brazilian and Thai children as well as Thai Adolescents (Burns et al. 2008; Burns et al. 2009). These studies also demonstrated convergent and discriminant validity for the ADHD-IN, ADHD-HI and ODD factors between mothers and fathers within each sample. In terms of the teacher CADBI, one study with Thai

adolescents and another with American children support the scale's construct validity (Shipp et al. 2010; Taylor et al. 2006) with one study demonstrating a scale specific correspondence between teacher ratings and direct observations of classroom behavior (Skansgaard and Burns 1998). No study has yet to examine the construct validity of the scale over mothers, fathers and teachers' ratings of the same adolescent.

Spanish Sample—ADHD Rating Scale-IV and the ODD Section of the Disruptive Behavior Disorders Rating Scale Mothers, fathers and teachers rated the occurrence of the nine ADHD-IN and ADHD-HI symptoms on a 4-point scale for the past 6-months (i.e., 0 = *never or rarely*; 1 = *sometimes*; 2 = *often*; and 3 = *very often*). The ADHD symptoms were rated on the ADHD Rating Scale-IV (DuPaul et al. 1998) with the ODD symptoms being rated with the ODD section of the Disruptive Behavior Disorders (DBD) rating scale (Barkley and Murphy 2005). Mothers and fathers were instructed to make their ratings independent of each other and the teachers had been interacting with the children for at least 8 months. The parents and teachers' ratings occurred at the same time.

Psychometric Properties The parent and teacher versions of the ADHD Rating Scale-IV are widely used measures of ADHD-IN and ADHD-HI symptoms. The scale has shown good internal consistency and four-week test-retest reliability for the ADHD-IN and ADHD-HI dimensions. The scale also predicts classroom behavior, task accuracy as well as diagnostic status and is sensitive to treatment effects (e.g., DuPaul et al. 1998). The ODD section of the DBD rating scale is also a widely used measure of the ODD symptoms with good reliability and validity (e.g., Barkley and Murphy 2005; Servera and Cardo 2007).

A recent study applied CFA to a *single* indicator by multitrait by multisource model with the Spanish sample (Servera et al. 2010). Rather than using the 26 symptoms as manifest variables, this study used total scores to represent each symptom dimension (e.g., the nine ADHD-IN symptoms are summed to create a single manifest variable for mothers). Although there was no evidence of convergent and discriminant validity between the mothers and teachers as well as fathers and teachers, evidence was found for convergent and discriminant validity between mothers and fathers. The amount of trait variance, however, varied substantially in the ADHD-IN, ADHD-HI and ODD manifest measures for mothers and fathers (i.e., 23 % to 64 %) with the source effects in the measures also being large for mothers and fathers (i.e., 29 % to 66 %) (Table 2, Servera et al. 2010). The single indicator model thus did not yield clear conclusions about the construct validity of the scale. Given the weaknesses of the single indicator model (see Eid

et al. 2006, p. 286–287 for a list of reasons the single indicator by multitrait by multisource model should not be used if multiple indicators are available), it was important to re-analyze the Spanish data with a multiple indicator model to determine if the multiple indicator model provides a clearer understanding of the scale's construct validity between mothers, fathers and teachers.

Analytic Strategy

A Multiple Indicator by Multitrait by Multisource Model Figure 1 shows the model for the analysis. The number of factors was a priori set to nine (i.e., ADHD-IN, ADHD-HI and ODD factors for mothers, fathers and teachers) with cross-loadings allowed *only* within each source. Each child had 78 symptom ratings (i.e., mothers provided 26 ratings, fathers 26 ratings and teachers 26 ratings). Correlated residuals were a priori specified between identical symptoms for mothers and fathers based on previous research (Burns et al. 2008, 2009).

Items as Categorical Indicators The symptom ratings were treated as ordered-categories. For the Thai sample, the seven and the eight-point anchors were collapsed into the six-point anchor due to the absence of ratings in the highest two categories for a few of the symptoms for teachers (the invariance analyses did not allow empty cells between sources).

Model Estimation Robust weighted least squares estimation (WLSMV) was used for the analyses (Mplus version 6.12, Muthén and Muthén 1998–2010). The multilevel modeling aspect of Mplus was also used to take account of each teacher rating multiple children (i.e., Type = Complex). All the models used Geomin rotation (Asparouhov and Muthén 2009). However, in order to calculate the amount of reliable variance in each factor, it was necessary to use robust maximum likelihood estimation (MLR) with the indicators treated as approximately continuous indicators. The calculation of the reliability coefficients was the only analysis that used robust maximum likelihood estimation.

Model Fit Three procedures were used to evaluate model fit. First, fit was evaluated with the comparative fit index (CFI, study criterion ≥ 0.95), Tucker Lewis Index (TLI, study criterion ≥ 0.95) and the root mean square error of approximation (RMSEA, study criterion ≤ 0.05). Second, localized ill fit was evaluated through an inspection of the residual matrix (i.e., ideally there should be no correlational residuals greater than an absolute value of 0.10, Kline 2011, p. 202). The third procedure involved an inspection of the model parameters.

Single Source Analyses The first set of analyses evaluated the fit of the ADHD-IN, ADHD-HI and ODD three-factor

Table 2 Invariance tests for ADHD-IN, ADHD-HI, and ODD factors between mothers, fathers and teachers' ratings

Invariance level	df	χ^2	CFI	TLI	RMSEA (90 % CI)
Thai Sample (n=872)					
1. Invariant form with correlated errors between like items for mothers' and fathers' ratings—baseline model	2725	2983	0.989	0.988	0.010 (0.007–0.013)
2. Invariant loadings and thresholds	3065	3329	0.989	0.989	0.010 (0.007–0.012)
Spanish sample (n=1749)					
1. Invariant form with correlated errors between like items for mothers' and fathers' ratings—baseline model.	2725	3631	0.989	0.988	0.014 (0.013–0.015)
2. Invariant loadings and thresholds	2961	3954	0.988	0.988	0.014 (0.013–0.015)

ADHD-IN attention-deficit/hyperactivity disorder—inattention, *ADHD-HI* attention-deficit/hyperactivity disorder—hyperactivity/impulsivity, *ODD* oppositional defiant disorder, *CFI* comparative fit index, *TLI* Tucker-Lewis index, *RMSEA* root mean-square error of approximation, *CI* confidence interval. The degrees of freedom for the invariant loadings and threshold model for the Thai sample are larger than for the invariant loadings and threshold model for the Spanish sample because there were five thresholds to constrain equal per item in the former compared to three thresholds per item in the later

model for each source separately. A good fit for each source was a necessary condition for the invariance analyses between mothers, fathers and teachers.

Invariance Analyses The baseline model did not constrain any parameters equal between mothers, fathers and teachers (other than constraints necessary for model identification). The next step involved constraining like-symptom loadings and thresholds equal between sources. The loadings and thresholds must be constrained simultaneously with categorical indicators (Muthén and Muthén 1998–2010, p. 433). Teachers were selected to be the reference group (i.e., the latent factor means for mothers and fathers were compared to teachers). For the baseline model, the variance of the

scale factors was set to one by default. For the model with the like-symptom loadings and thresholds constrained equal, the scale factors for the teacher ratings were set to one because teachers were the reference group with the scale factors for the mothers and fathers' ratings being estimated (L. K. Muthén, personal communication, November 1, 2009).

Two different procedures were used to determine if the model with like-symptom loadings and thresholds constrained equal between sources was equivalent to the baseline model. The first procedure used changes in the CFI, TLI, and RMSEA from the baseline model to the model with the constraints. A decrease in the CFI of ≤ 0.01 (Chen 2007) in conjunction with no change (or an improvement) in

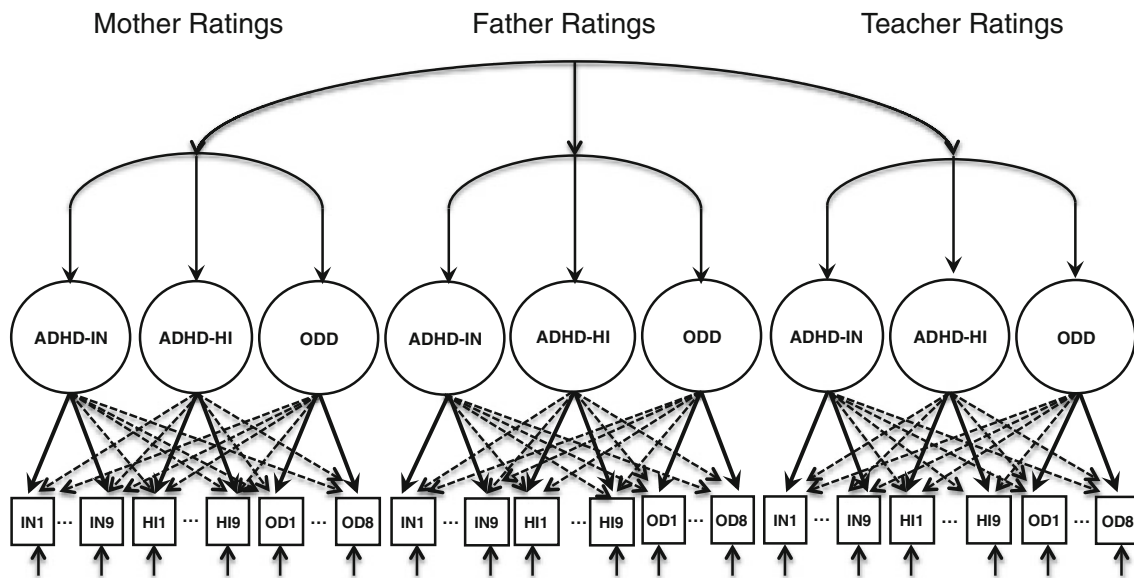


Fig. 1 Baseline model used for the application of exploratory structural equation modeling to a multiple indicator (26 individual symptoms) by multitrait (ADHD-IN, ADHD-HI and ODD) by multisource (mothers, fathers and teachers) model. Cross-loadings (dashed lines)

were only allowed within each source. Although not shown in the model, correlated residuals were included between the same symptoms for mothers and fathers' ratings

the TLI and RMSEA was used to suggest no meaningful decrement in model fit. The second procedure compared the residuals from the baseline model to the residuals for the model with the loadings and thresholds constrained equal. If the model with the constraints did not result in greater localized ill fit than the baseline model (e.g., number of residuals greater than 0.10), then such an outcome was also used to suggest no meaningful decrement in model fit.

Convergent Validity, Discriminant Validity and Source Effects Convergent validity *between* sources required the same factor-different source correlations to be substantial (e.g., 0.70) while discriminant validity *between* sources required the same factor-different source correlations to be significantly larger than the different factor-different source correlations. Discriminant validity *within* sources required the ADHD-IN, ADHD-HI and ODD factor correlations to correlate no higher than 0.80 to 0.85 (Brown 2006, p. 131). The magnitude of the source effects was estimated by a comparison of the different factor-same source correlations to the different factor-different source correlations (Brown 2006, chap. 6). Source effects are present to the extent that the different factor-same source correlations are larger than the different factor-different source correlations (i.e., To what extent does a common source inflate the correlation between the ADHD-IN and ADHD-HI factors relative to the correlation between the ADHD-IN and ADHD-HI factors based on different sources?).

Missing Data Procedure Each adolescent in the Thai sample had a mother, father, and teacher rating. For the Spanish sample, some of the children were missing a mother's rating (i.e., 1,422 of the 1,749 children were rated by mothers) and some of the children were missing a father's rating (i.e., 1,380 of the 1,749 children were rated by fathers) while no children were missing a teacher's rating. WLSMV estimation uses pairwise deletion to deal with the missing information.

Results

Fit of ADHD-IN, ADHD-HI, and ODD Three-Factor Model for Mothers, Fathers, and Teachers Separately

Thai Adolescents Table 1 shows the fit indices for the three-factor model for mothers, fathers and teachers separately for the Thai sample. All the CFI and TLI values were ≥ 0.974 and all the RMSEA values were ≤ 0.039 . For mothers' ratings, only 2 of the 325 correlation residuals were larger than an absolute value of 0.10 (values of 0.109 and 0.136), for fathers' ratings only 3 (values of 0.105, 0.118, and 0.133), and for teachers' ratings none. The three-factor model thus resulted in a good fit for each source separately for the Thai adolescents.

Spanish Children The three-factor model also resulted in good fit for each source for the Spanish sample. All the CFI and TLI values were ≥ 0.963 and all the RMSEA values were ≤ 0.047 (Table 1). For mothers' ratings, only 4 of the 325 correlation residuals were larger than an absolute value of 0.10 (range = 0.109 to 0.117), for fathers' ratings only 5 (range = 0.115 to 0.127), and for teachers' ratings none.

Invariance of the ADHD-IN, ADHD-HI, and ODD Model Between Mothers, Fathers and Teachers

Thai Adolescents Table 2 shows the fit for the baseline model and the model with the like-symptom loadings and thresholds constrained equal for the Thai adolescents. The baseline model resulted in a good fit (i.e., CFI = 0.989, TLI = 0.988, and RMSEA = 0.010) with no major areas of localized ill fit (i.e., only 38 of the 3003 residuals were greater than 0.10, range = 0.101 to 0.150). The model with like-symptom loadings and thresholds constrained equal also resulted in a good global fit (i.e., CFI = 0.989, TLI = 0.989, and RMSEA = 0.010) with no meaningful decrement in fit from the baseline model (i.e., CFI unchanged, TLI slightly better, and RMSEA unchanged) and no increase in localized ill fit (i.e., only 38 of 3003 correlation residuals were greater than 0.10, range = 0.101 to 0.150).

Spanish Children Table 2 shows the fit for the baseline mode as well as the model with the like-symptom loadings and thresholds constrained equal for the Spanish sample. The baseline model resulted in a good fit (i.e., CFI = 0.989, TLI = 0.988 and RMSEA = 0.014) and no major areas of localized ill fit (i.e., only 20 of the 3003 correlation residuals were greater than 0.10, range = 0.101 to 0.150). The model with like-item loadings and thresholds constrained equal also resulted in a good fit (i.e., CFI = 0.988, TLI = 0.988 and RMSEA = 0.014) with the decrease in the CFI value being only 0.001 and the TLI and RMSEA being unchanged. For the 3003 correlation residuals, 23 were greater than 0.10 (range = 0.101 to 0.150). The model with the constraints did not result in a meaningful decrement in fit or an increase in localized ill fit from the baseline model for either the Thai or Spanish samples.

ADHD-IN, ADHD-HI and ODD Symptom-Factor Loadings

Table 3 shows the primary and secondary loadings of the symptoms on the three factors. Only one set of primary and secondary loading is shown for each sample given the invariance of like-symptoms loadings between mothers, fathers and teachers within each sample. Ideally a symptom's loadings on its primary factor should be substantial (e.g., ≥ 0.70) with its loadings on the two secondary factors being close to zero (e.g., $\leq \pm 0.20$).

Table 3 Standardized (probit coefficients) symptom-factor loadings (standard errors)

ADHD-IN Symptoms	Thai sample (n=872)			Spanish sample (n=1749)		
	ADHD-IN	ADHD-HI	ODD	ADHD-IN	ADHD-HI	ODD
1. Close Attention	0.88 (0.03)	0.00 (0.02) ^{ns}	0.04 (0.05) ^{ns}	0.92 (0.03)	0.12 (0.03)	-0.09 (0.03)
2. Sustaining Attention	0.90 (0.03)	0.01 (0.02) ^{ns}	0.05 (0.04) ^{ns}	0.72 (0.03)	0.30 (0.03)	-0.03 (0.02) ^{ns}
3. Listen	0.63 (0.04)	0.09 (0.05) ^{ns}	0.24 (0.04)	0.52 (0.03)	0.33 (0.03)	0.15 (0.03)
4. Follow Through	0.80 (0.04)	-0.06 (0.04) ^{ns}	0.21 (0.04)	0.88 (0.02)	0.07 (0.03)	0.03 (0.03) ^{ns}
5. Organizational Skills	0.83 (0.04)	-0.03 (0.04) ^{ns}	0.14 (0.05)	0.93 (0.02)	0.00 (0.02) ^{ns}	0.02 (0.03) ^{ns}
6. Concentration	0.85 (0.04)	0.05 (0.05) ^{ns}	0.08 (0.05) ^{ns}	0.86 (0.03)	-0.03 (0.02) ^{ns}	0.10 (0.03)
7. Loses Things	0.74 (0.06)	0.21 (0.06)	0.01 (0.02) ^{ns}	0.70 (0.03)	0.13 (0.03)	0.12 (0.04)
8. Easily Distracted	0.65 (0.02)	0.33 (0.05)	-0.01 (0.03) ^{ns}	0.72 (0.02)	0.26 (0.03)	-0.01 (0.01) ^{ns}
9. Forgetful	0.76 (0.07)	0.31 (0.06)	-0.13 (0.05)	0.84 (0.03)	-0.05 (0.04) ^{ns}	0.14 (0.04)
ADHD-HI Symptoms						
1. Fidgets/Squirms	0.05 (0.04) ^{ns}	0.80 (0.05)	0.01 (0.03) ^{ns}	0.15 (0.03)	0.82 (0.03)	-0.01 (0.02) ^{ns}
2. Leaves Seat	-0.01 (0.03) ^{ns}	0.86 (0.05)	0.04 (0.04) ^{ns}	0.22 (0.03)	0.77 (0.02)	0.00 (0.01) ^{ns}
3. Runs/Climbs	-0.07 (0.06) ^{ns}	0.99 (0.05)	-0.07 (0.08) ^{ns}	0.16 (0.03)	0.61 (0.03)	0.25 (0.04)
4. Playing Quietly	0.02 (0.03) ^{ns}	0.90 (0.05)	-0.04 (0.04) ^{ns}	0.39 (0.03)	0.35 (0.03)	0.27 (0.04)
5. Talks Excessively	-0.04 (0.03) ^{ns}	0.85 (0.06)	0.09 (0.05) ^{ns}	0.00 (0.02) ^{ns}	0.90 (0.03)	0.06 (0.03)
6. Driven/On the Go	-0.16 (0.05)	0.94 (0.06)	0.04 (0.04) ^{ns}	-0.05 (0.02)	0.76 (0.03)	0.16 (0.03)
7. Blurts	0.08 (0.04)	0.74 (0.05)	0.12 (0.03)	-0.06 (0.02)	0.64 (0.03)	0.37 (0.04)
8. Awaiting Turn	0.12 (0.05)	0.67 (0.05)	0.12 (0.03)	0.01 (0.02) ^{ns}	0.69 (0.03)	0.28 (0.03)
9. Interrupts/Intrudes	0.06 (0.05) ^{ns}	0.85 (0.04)	-0.01 (0.03) ^{ns}	0.01 (0.02) ^{ns}	0.61 (0.03)	0.38 (0.03)
ODD Symptoms						
1. Argues	0.00 (0.01) ^{ns}	0.05 (0.04) ^{ns}	0.88 (0.04)	0.01 (0.02) ^{ns}	0.18 (0.04)	0.80 (0.02)
2. Loses Temper	-0.04 (0.04) ^{ns}	-0.07 (0.04) ^{ns}	0.99 (0.02)	-0.08 (0.03)	0.25 (0.04)	0.80 (0.03)
3. Refuses	0.27 (0.03)	0.02 (0.03) ^{ns}	0.65 (0.03)	0.07 (0.03)	0.17 (0.04)	0.77 (0.03)
4. Annoys	0.13 (0.05)	0.22 (0.05)	0.63 (0.04)	0.07 (0.03)	0.29 (0.03)	0.66 (0.03)
5. Blames	0.05 (0.04) ^{ns}	0.11 (0.05)	0.79 (0.03)	0.05 (0.02)	0.21 (0.03)	0.71 (0.03)
6. Gets Annoyed	-0.08 (0.04)	0.02 (0.03) ^{ns}	0.95 (0.03)	-0.03 (0.02) ^{ns}	-0.02 (0.04) ^{ns}	0.95 (0.03)
7. Resentful	-0.04 (0.03) ^{ns}	0.00 (0.02) ^{ns}	0.96 (0.03)	0.00 (0.01) ^{ns}	-0.15 (0.04)	0.99 (0.02)
8. Spiteful/Vindictive	0.03 (0.05) ^{ns}	0.23 (0.06)	0.65 (0.05)	-0.07 (0.03)	0.00 (0.01) ^{ns}	0.93 (0.02)

There is only one set of loadings for each sample because the loadings were invariant over mothers, fathers and teachers' ratings within each sample. Primary loadings less than 0.70 and secondary loadings greater than 0.20 are in bold to note symptoms with weaker discriminant validity. All loadings significant at $p < 0.05$ unless indicate as ^{ns}

Thai Sample For the Thai sample, 20 of the 26 primary loadings were greater than 0.69 with the six symptoms that did not meet the 0.70 criteria having primary loadings from 0.63 to 0.67. For the 52 secondary loadings, 8 symptoms had loadings greater than 0.20 (range = 0.21 to 0.33) with most of the cross-loadings being close to zero (e.g., 34 of 52 did not differ significantly from 0.00). The six symptoms with the weakest discriminant validity were the ADHD-IN symptoms “listen,” “easily distracted” and “forgetful” along with the ODD symptoms “refuses,” “annoys” and “spiteful/vindictive.”

Spanish Sample For the Spanish sample 19 of the 26 primary loadings were greater than 0.69 with the range for the symptoms with primary loading less than 0.69 being 0.35 to 0.69. For the 52 secondary loadings, 13 symptoms had secondary

loadings greater than 0.20 (range = 0.21 to 0.39) with 17 of the 52 secondary loading not being statistically different from 0.00. The ADHD-HI symptom “playing quietly” showed no discriminant validity (i.e., higher loading on the ADHD-IN factor than on the ADHD-HI factor) with ADHD-IN symptoms “sustaining attention,” “listen,” and “easily distracted,” the ADHD-HI symptoms “runs/climbs,” “blurts,” “awaiting turn” and “interrupts/intrudes” and the ODD symptom “annoys” having weak discriminant validity.

Latent Mean Differences on the ADHD-IN, ADHD-HI and ODD Factors Between Sources

Table 4 shows latent mean results for the Thai adolescents and Spanish children. For the Thai sample, mothers and fathers rated

Table 4 Latent mean differences for mothers' and fathers' ratings relative to teachers' ratings for the ADHD-IN, ADHD-HI and ODD factors

Factors	Thai adolescents ($n=872$)			Spanish children ($n=1749$)		
	Difference (SE)	z	d	Difference (SE)	z	d
ADHD-IN						
Teachers	0.00 (–)	–	–	0.00 (–)	–	–
Mothers	0.08 (0.13)	0.63 ^{ns}	0.12	0.13 (0.06)	2.40*	0.18
Fathers	0.10 (0.13)	0.80 ^{ns}	0.15	0.14 (0.05)	2.65*	0.21
ADHD-HI						
Teachers	0.00 (–)	–	–	0.00 (–)	–	–
Mothers	0.50 (0.14)	3.62*	0.71	0.49 (0.06)	8.34*	0.80
Fathers	0.52 (0.15)	3.57*	0.79	0.51 (0.06)	8.54*	0.83
ODD						
Teachers	0.00 (–)	–	–	0.00 (–)	–	–
Mothers	1.01 (0.14)	7.34*	1.63	0.72 (0.06)	12.16*	1.25
Fathers	0.98 (0.14)	7.24*	1.69	0.67 (0.06)	11.09*	1.20

Teachers are the reference group with the latent factor means for mothers and fathers compared to teachers. The mean difference for the Thai sample is based on a 6-point rating scale and the mean difference for the Spanish sample is based a 4-point rating scale. d = mean difference in standard deviation units. * $p < 0.05$

the adolescents as significantly ($ps < 0.001$) higher than teachers on the ADHD-HI and ODD factors with the difference being non-significant for the ADHD-IN factor. For the Spanish sample, mothers and fathers rated the children as significantly higher on the ADHD-IN, ADHD-HI and ODD factors than teachers ($ps < 0.001$). Mothers and fathers' ratings were almost the same amount higher than teachers in both samples.

Reliability and Discriminant Validity of ADHD-IN, ADHD-HI and ODD Within Sources

Thai Adolescents Table 5 shows the correlations among the ADHD-IN, ADHD-HI and ODD factors within sources along with the reliability coefficients for the Thai adolescent. The reliability coefficient for each dimension was good (0.95 to 0.98). In addition, the within source factor correlations ranged from 0.53 to 0.77, thus indicating the three factors met the minimum criteria for discriminant validity within each source (Brown 2006, p. 131).

Spanish Children Table 6 shows the correlations among the ADHD-IN, ADHD-HI and ODD factors within sources along with the reliability coefficients for the Spanish children. The reliability coefficients were good (0.92 to 0.98) with the within source factor correlations ranging from 0.43 to 0.62. The three factors thus again met the minimum criteria for discriminant validity within each source (i.e., all the within source factor correlations were less than 0.85, Brown 2006, p. 131).

Convergent and Discriminant Validity for ADHD-IN, ADHD-HI, and ODD Between Sources

Thai Adolescents Table 5 shows the convergent and discriminant validity correlations for the three factors between mothers, fathers and teachers for the Thai adolescents. For mothers and fathers, each convergent validity correlation was significant ($ps < 0.001$) and substantial (i.e., $M = 0.70$, range = 0.67 to 0.73) with the convergent correlations being significantly larger ($ps < 0.001$) than the mother-father discriminant correlations (i.e., $M = 0.46$, range = 0.38 to 0.52). There was no meaningful convergent validity between mothers and teachers as well as fathers and teachers.

Spanish Children Table 6 shows the convergent and discriminant validity correlations for the three factors between mothers, fathers and teachers for the Spanish children. For mothers and fathers, each of the convergent validity coefficients was significant ($ps < 0.001$) and substantial (i.e., $M = 0.86$, range = 0.78 to 0.91) with the convergent correlations being significantly ($ps < 0.001$) larger than the mother-father discriminant correlations (i.e., $M = 0.42$, range = 0.39 to 0.45). For mothers and teachers as well as fathers and teachers, there was moderate support for the convergent and discriminant validity of the ADHD-IN and ADHD-HI factors. For the mother-teacher correlations, the convergent correlations for ADHD-IN and ADHD-HI factors were 0.62 and 0.50 ($ps < 0.001$), respectively, with these values being significantly ($ps < 0.001$) larger than the discriminant correlations (i.e., $M = 0.23$, range = 0.13 to 0.32). Almost identical results occurred for the father-teacher ADHD-IN and ADHD-HI convergent and discriminant correlations. The ODD factor, however, failed to show convergent and discriminant validity for the parent-teacher comparisons.

Source Effects for the ADHD-IN, ADHD-HI, and ODD Factors

Thai Adolescents The mean of the different factor-different source correlations for mothers and fathers was 0.46 (range = 0.38 to 0.52) with the mean of the different factor-same source correlations being 0.63 (range = 0.53 to 0.71). The variance associated with the within source correlations was approximately 19 % larger than variance associated with the between source correlations, thus indicating moderate source effects. Given the lack of convergent and discriminant validity between parents and teachers, the calculation of the source effects was not meaningful.

Spanish Children The mean of the different factor-different source correlations for mothers and fathers was 0.42 (range = 0.39 to 0.45) with the mean of the different factor-same source correlations being 0.46 (range = 0.44

Table 5 Multitrait by multi-source latent factor correlations for Thai adolescents

	Mothers			Fathers			Teachers		
	IN	HI	ODD	IN	HI	ODD	IN	HI	ODD
Mothers									
ADHD-IN	(0.95)								
ADHD-HI	0.71	(0.96)							
ODD	0.53	0.59	(0.96)						
Fathers									
ADHD-IN	0.70	0.51	0.38	(0.95)					
ADHD-HI	0.52	0.73	0.48	0.71	(0.95)				
ODD	0.40	0.44	0.67	0.56	0.67	(0.96)			
Teachers									
ADHD-IN	0.26	0.13	0.15	0.27	0.15	0.15	(0.98)		
ADHD-HI	0.12	0.13	0.05 ^{ns}	0.09 ^{ns}	0.09^{ns}	.09 ^{ns}	0.60	(0.98)	
ODD	0.06 ^{ns}	0.09	0.11	0.09 ^{ns}	0.11 ^{ns}	0.11^{ns}	0.61	0.77	(0.97)

Reliability coefficients are on the main diagonal in parentheses. Convergent validity correlations are the values in bold type. All correlations are significant at $p < 0.05$ unless indicated as ^{ns}. $N = 872$

to 0.49). The variance associated with the within source correlations was approximately 3 % larger than the variance associated with the between source correlations, thus indicating no meaningful source effects for mothers and fathers. The average for the different factor-different source correlations for mothers and teachers for the ADHD-IN and ADHD-HI factors was 0.26 (range 0.13 to 0.36) with the average of the different factor-same source correlations being 0.52. Almost identical results occurred for the father-teacher comparisons for the ADHD-IN and ADHD-HI factors. Here the source effects were larger than for the mother-father comparisons due to the within source correlations for teachers being larger than the within source correlations for mothers and fathers (i.e., the variance associated with the within source correlations was approximately 20 % larger than the variance associated with the between source correlations).

Discussion

Our purpose was to demonstrate the usefulness of exploratory SEM with multiple sources for the evaluation of the construct validity of ADHD/ODD rating scales. With multiple sources the exploratory SEM procedure can determine the (1) convergent and discriminant validity of the individual symptoms for each source; (2) equality of like symptom loadings and thresholds between sources; (3) equality of like factor means between sources; (4) convergent and discriminant validity of the factors between sources; (5) discriminant validity of the factors within sources; and (6) magnitude of source effects. We will summarize our findings for the two DSM-IV scales in the context of the six questions and then outline the usefulness of the procedure for the evaluation of the forthcoming DSM-V ADHD/ODD scales.

Table 6 Multitrait by multi-source latent factor correlations for Spanish children

	Mothers			Fathers			Teachers		
	IN	HI	ODD	IN	HI	ODD	IN	HI	ODD
Mothers									
ADHD-IN	(0.95)								
ADHD-HI	0.46	(0.94)							
ODD	0.46	0.44	(0.93)						
Fathers									
ADHD-IN	0.91	0.45	0.43	(0.95)					
ADHD-HI	0.44	0.89	0.41	0.49	(0.94)				
ODD	0.41	0.39	0.78	0.49	0.43	(0.92)			
Teachers									
ADHD-IN	0.62	0.32	0.18	0.60	0.29	0.25	(0.98)		
ADHD-HI	0.27	0.50	0.13	0.29	0.45	0.13	0.59	(0.98)	
ODD	0.27	0.36	0.26	0.30	0.35	0.30	0.52	0.62	(0.97)

Reliability coefficients are on the main diagonal in parentheses. Convergent validity correlations are the values in bold type. All correlations are significant at $p < 0.05$. $N = 1749$

Convergent and Discriminant Validity of Individual Symptoms Within Sources

Convergent validity requires the loadings of the symptoms on their primary factor to be substantial while discriminant validity requires the loadings on the primary factor to be larger than the loadings on secondary factors (e.g., primary loadings ≥ 0.70 and secondary loadings $\leq \pm 0.20$). The results with the CADBI for the Thai sample indicated that all the symptoms showed reasonable to good convergent validity as well as most of the symptoms showing reasonable to good discriminant validity. For the Spanish children with the DSM-IV ADHD scale and the ODD scale from the DBD, the convergent and discriminant validity results were not as strong (i.e., the convergent and discriminant validity of the symptoms was weaker and one ADHD-HI symptom showed no discriminant validity). There were, however, three consistent findings across the two samples. The results indicated that the ADHD-IN symptoms “listen” and “easily distracted” as well as the ODD symptom “annoys” had weak discriminant validity.

Additional work on the wording of these symptoms might improve their content validity and thus their construct validity (Shipp et al. 2010, pp. 558–560). The most important point here, however, is that the magnitudes of some of the secondary loadings, especially for the Spanish sample, indicate that it would be inappropriate to restrict all the secondary loadings to zero. The secondary loadings in Table 3 make a strong case for the use of exploratory SEM for the initial evaluation of the forthcoming DSM-V ADHD/ODD scales.

Equality of Symptom Loadings, Symptom Thresholds and Factor Means Between Sources

The like-symptom loadings and thresholds for the Thai adolescents and Spanish children were invariant between sources, thus indicating these aspects of the measurement structure of the ADHD-IN, ADHD-HI and ODD factors did not change over mothers, fathers and teachers. The demonstration of the invariance of symptom loadings and thresholds then allowed an evaluation of the invariance of the factor means between sources. In the Spanish sample, mothers and fathers rated the children significantly higher on the ADHD-IN, ADHD-HI and ODD factors than teachers with mothers and fathers’ ratings being equally higher than teachers). Similar results occurred in the Thai sample. ODD also had a much higher occurrence in the home than in the school (i.e., parents’ ratings of ODD were approximately 1.23 and 1.66 standard deviations higher than teachers for the Spanish and Thai samples, respectively—perhaps due to the school environment being less tolerant of ODD behavior than the home environment or children and adolescents being more respectful of teachers than parents).

Convergent and Discriminant Validity of ADHD-IN, ADHD-HI and ODD Between Sources

For the Thai and Spanish samples, the ADHD-IN, ADHD-HI and ODD factors showed statistically significant convergent and discriminant validity between mothers and fathers. In addition, for the Spanish sample, the ADHD-IN and ADHD-HI factors showed statistically significant convergent and discriminant validity between mothers and teachers as well as fathers and teachers (although less strong than for mothers and fathers). For the Thai sample, however, the ADHD-IN and ADHD-HI factors did not show convergent validity between parents and teachers. This difference between the two samples could be due to several factors (e.g., the different ages, the use of different scales or the different cultures). There was also no support for the convergent validity of the ODD factor between parents and teachers for either sample, thus indicating no meaningful relative stability of the children’s and adolescent’s behavior on this factor between home and school. A future study will need to include two sources in the school in addition to two sources in the home to clarify the convergent and discriminant validity results between home and school. For example, if mothers and fathers as well as teachers and aides show strong convergent and discriminant within home and school, respectively, then such findings would provide a better foundation to study the home to school validity question.

Discriminant Validity of ADHD-IN, ADHD-HI and ODD Within Sources

In order to establish the external validity of the ADHD-IN, ADHD-HI and ODD factors (e.g., the identification of unique causes, associated features, risk factors, outcomes), each factor must contain enough unique variance (i.e., variance independent of the other two factors) to allow for the possibility of the identification of unique correlates of each factor. The correlations among the three factors within sources ranged from 0.43 to 0.77, thus indicating enough independence for the symptom dimensions within a source for a meaningful search for unique external correlates for each dimension. It should also be noted that the ADHD-IN, ADHD-HI and ODD factor correlations were lower than the factor correlations reported in CFA studies. This result was due to the use of the exploratory SEM procedure that does not restrict the cross-loadings to zero (Asparouhov and Muthén 2009).

Magnitude of Source Effects

Although the use of exploratory CFA to model a multiple indicator by multitrait by multisource matrix does not allow for the calculation of source effects as a latent variable (see

weaknesses section below), this procedure does allow for an estimation of the source effects (i.e., Are the different factor-same source correlations larger than the different factor-different source correlations?). For the Thai sample, the within source correlations were larger than the between source correlations for mothers and fathers (i.e., the variance associated with the within source correlations was approximately 19 % larger than the variance associated with the between source correlations), thus indicating the presence of source effects. For the Spanish sample, however, there were no meaningful source effects for mothers and fathers ratings (i.e., the variance associated with the within source correlations was approximately 3 % more than the between source correlations). Source effects did occur for the mothers and teachers as well as fathers and teachers for the Spanish sample for the ADHD-IN and ADHD-HI factors (i.e., the variance associated with the within source correlations was approximately 20 % more than the between source correlations).

Summary of Findings

Most the individual ADHD/ODD symptoms showed convergent and discriminant validity, especially for the CADBI. The ADHD-IN, ADHD-HI and ODD latent factors also showed convergent and discriminant validity *between* mothers and fathers for both scales with only the ADHD-IV Rating Scale demonstrating convergent and discriminant validity for the ADHD-IN and ADHD-HI factors between parents and teachers. In addition, the three latent factors showed discriminant validity *within* mothers, fathers and teachers for both scales. Source effects, however, were stronger for the CADBI than in the other scale, especially for mothers and fathers. The findings also indicated that exploratory SEM was a more appropriate procedure to apply to the data than CFA, especially for the scale used in the Spanish sample.

A Multiple Indicator by Multitrait by Multisource Model Versus Single Indicator by Multitrait by Multisource Model—Different Findings

The use of a multiple indicator by multitrait by multisource model in current study resulted in two different outcomes from the single indicator by multitrait by multisource model with the Spanish sample in the earlier study (Servera et al. 2010). First, a multiple indicator model found minimum source effects for mothers and fathers while the single indicator model found much larger source effects. Second, a multiple indicator model found support for the convergent and discriminant validity of the ADHD-IN and ADHD-HI factors between mothers and teachers as well as fathers and teachers while the single indicator model did not find such convergent and discriminant validity. These two different

outcomes along with the more comprehensive and clearer results support Eid et al.'s (2006, p. 292) argument that a single indicator model should not be used if multiple indicators are available for each trait-source unit.

Weaknesses of a Multiple Indicator by Multitrait by Multisource Model

The major weakness of a multiple indicator by multitrait by multisource model is the inability to separate the variability in the individual symptom ratings into *latent* source and *latent* trait effects (e.g., How much of the variance in the symptom ratings for mothers, fathers and teachers is trait variance, source variance and residual?). If the research question requires latent source and trait factors in order to relate these factors to predictors and outcomes, then a multiple indicator by correlated trait by correlated method minus one model represents a better choice than the multiple indicator model of the current study. Eid and colleagues have described the usefulness of a multiple indicator by correlated trait by correlated method minus one model for examination of trait and source effects (Eid et al. 2006). Dumenci et al. (2011) also presented a novel model to measure context specific and cross-contextual effects in multiple source rating scales. Researchers now have an increasing number of sophisticated multiple indicator models to study different aspects of the construct validity of multisource rating scales.

Recommendations for the Development of DSM-V ADHD/ODD Rating Scales

Given the important role of ADHD/ODD rating scales in research on these disorders (e.g., identification of associated features, risk factors and outcomes), the forthcoming DSM-V rating scales need to have the best construct validity possible. In our opinion, the minimum conditions for the scales' use in research and clinical activities should include positive answers to our six questions for two sources *within the same situation* (e.g., mothers and fathers in the home; teachers and aides in the school) with two occasions of measurement (i.e., a short test-retest interval). With two occasions of measurement added to the model, the six questions could be evaluated across the two assessment occasions as well as allowing for an assessment of the convergent and discriminant validity of *change* on the three factors (Geiser et al. 2010).

Answers to the six questions for two sources in the home as well as two sources within the school would provide the foundation for a more meaningful study of the more complex aspects of construct validity for ADHD/ODD scales (e.g., the question of construct validity between home and school, the study of construct validity with the more

complex models of Eid and Dumenci) as well as a more fruitful search for external correlates of the factors. We encourage researchers to apply exploratory SEM to a multiple indicator by multitrait by multisource by multioccasion model to understand better the construct validity of the DSM-V ADHD/ODD scales. This procedure has the ability to identify ADHD/ODD symptoms with weak or no discriminant validity in the scales. Such information would provide the basis to improve the content validity of the symptoms in the scale as well as identify symptoms that do not have enough discriminant validity to belong with the symptom dimension.

References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438.
- Barkley, R. A. (2011). Distinguishing Sluggish Cognitive Tempo from Attention-Deficit Hyperactivity Disorder in adults. *Journal of Abnormal Psychology*. Advance online publication. doi:10.1037/a0023961.
- Barkley, R. A., & Murphy, K. R. (2005). *Attention-deficit hyperactivity disorder: A clinical workbook* (3rd ed.). New York: Guilford.
- Bauermeister, J. J., Canino, G., Polanczyk, G., & Rohde, L. A. (2010). ADHD across cultures: is there evidence for a bidimensional organization of symptoms? *Journal of Clinical Child and Adolescent Psychology*, *39*, 362–372.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Burns, G. L., & Haynes, S. N. (2006). Clinical psychology: Construct validation with multiple sources of information and multiple settings. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association.
- Burns, G. L., Taylor, T., & Rusby, J. (2001a). Child and adolescent disruptive behavior inventory—Version 2.3 (Parent). Unpublished manuscript.
- Burns, G. L., Taylor, T., & Rusby, J. (2001b). Child and adolescent disruptive behavior inventory—Version 2.3 (Teacher). Unpublished manuscript.
- Burns, G. L., Walsh, J. A., Boe, B., Sommers-Flanagan, R., & Teegarden, L. A. (2001c). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: what is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology*, *29*, 339–349.
- Burns, G. L., Moura, M. A., Walsh, J. A., Desmul, C., Silpakit, C., & Sommers-Flanagan, J. (2008). Invariance and convergent and discriminant validity between mothers' and fathers' ratings of oppositional defiant disorder toward adults, ADHD-HI, ADHD-IN, and academic competence factors within Brazilian, Thai, and American children. *Psychological Assessment*, *20*, 121–130.
- Burns, G. L., Desmul, C., Walsh, J. A., Silpakit, C., & Ussahawanitchakit, P. (2009). A multitrait (ADHD-IN, ADHD-HI, ODD toward adults, academic and social competence) by multisource (mothers and fathers) evaluation of the invariance and convergent/discriminant validity of the Child and Adolescent Disruptive Behavior Inventory with Thai Adolescents. *Psychological Assessment*, *21*, 635–641.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504.
- Dumenci, L., Achenbach, T. M., & Windle, M. (2011). Measuring context-specific and cross-contextual components of hierarchical constructs. *Journal of Psychopathology and Behavioral Assessment*, *33*, 3–10.
- DuPaul, G. J., Power, T. J., McGoey, K. E., Ikeda, M. J., & Anastopoulos, A. D. (1998). Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *Journal of Psychoeducational Assessment*, *16*, 55–68.
- Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation modeling for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, DC: American Psychological Association.
- Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: application of a multimethod change model to depression and anxiety in children. *Developmental Psychology*, *46*, 29–45.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford.
- Moura, M. A., & Burns, G. L. (2010). Oppositional defiant disorder toward adults and oppositional defiant disorder toward other children: evidence for two separate constructs with mothers and fathers' ratings of Brazilian children. *Journal of Child Psychology and Psychiatry*, *51*, 23–30.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus User's Guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Servera, M., & Cardo, E. (2007). Las ADHD Rating Scale-IV en una muestra escolar española: datos normativos y consistencia interna para maestros, padres y madres (*ADHD Rating Scale-IV in a sample of Spanish schoolchildren: normative data and internal consistency of teacher and parents*). *Revista de Neurología*, *45*(7), 393–399.
- Servera, M., Lorenzo-Seva, U., Cardo, E., Rodríguez-Fornells, A., & Burns, G. L. (2010). Understanding trait and source effects in Attention Deficit Hyperactivity Disorder and Oppositional Defiant Disorder rating scales: mothers, fathers and teachers' ratings of children from the Balearic Islands. *Journal of Clinical Child and Adolescent Psychology*, *39*, 1–11.
- Shipp, F., Burns, G. L., & Desmul, C. (2010). Construct validity of ADHD-IN, ADHD-HI, ODD toward adults, academic and social competence dimensions in teacher ratings of Thai Adolescents. *Journal of Psychopathology and Behavioral Assessment*, *32*, 557–564.
- Skansgaard, E. P., & Burns, G. L. (1998). Comparison of DSM-IV ADHD combined type and predominantly inattentive types: correspondence between teacher ratings and direct observations of inattention, hyperactivity/impulsivity, slow cognitive tempo, oppositional defiant disorder, and overt conduct disorder symptoms. *Child and Family Behavior Therapy*, *20*, 1–14.
- Taylor, T. K., Burns, G. L., Rusby, J. C., & Foster, E. M. (2006). Oppositional defiant disorder toward adults and oppositional defiant disorder toward peers: Initial evidence for two separate constructs. *Psychological Assessment*, *18*, 439–433.
- Toplak, M. E., Pitch, A., Flora, D. B., Iwenofu, L., Ghelani, K., Jain, U., & Tannock, R. (2009). The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: evidence for a general factor with separable dimensions. *Journal of Abnormal Child Psychology*, *37*, 1137–1150.
- Toplak, M. E., Sorge, G. B., Flora, D. B., Chen, W., Banaschewski, T., Buitelaar, J., et al. (2012). The hierarchical factor model of ADHD: invariant across age and national groupings? *Journal of Child Psychology and Psychiatry*, *53*, 292–303.