

## Federation University ResearchOnline

<https://researchonline.federation.edu.au>

Copyright Notice

This is the peer-reviewed version of the following article:

Gomez, Vance, A., & Stavropoulos, V. (2020). Correlated Trait–Correlated Method Minus One Analysis of the Convergent and Discriminant Validity of the Conners 3 Short Forms. *Assessment (Odessa, Fla.)*, 27(7), 1463–1475.


The online version of this article can be found at:

<https://doi.org/10.1177/1073191118803714>

Copyright © The Author(s) 2018 who receive access to an article through a repository are reminded that the article is protected by copyright. Users may download and save a local copy of an article accessed in an institutional repository for the user's personal reference. For permission to reuse an article, please follow the SAGE [Process for Requesting Permission](#).

See this record in Federation ResearchOnline at:

<http://researchonline.federation.edu.au/vital/access/HandleResolver/1959.17/174045>

Rapson Gomez<sup>1,2</sup>, Alasdair Vance<sup>3,4</sup>, and Vasileios Stavropoulos<sup>1,5</sup> 

### Abstract

This study used the correlated trait–correlated method minus one model to examine the convergent and discriminant validity of the scales of the Conners 3 Short [C 3 (S)]. The C 3 (S) scales in the analysis were inattention (IN), hyperactivity/impulsivity (HY), learning problems (LP; learning problems/executive functioning from the teacher version), aggression (AG), and peer relations (PR, only for parent and teacher versions). A total of 529 adolescents and children (75% males, mean age = 11.75 years,  $SD = 2.97$  years) provided self-ratings, and were also rated by their mothers and teachers. The findings indicated no support for the convergence of IN and HY across the three respondents. In contrast, there was convergence for LP, AG, and PR. There was support for the discriminant validity of the traits, except between IN and HY. The findings are discussed in relation to the convergent and discriminant validity of the C 3 (S) measures, and the clinical use of the C 3 (S).

### Keywords

Conners 3 Short, correlated trait–correlated method minus one model, mother, teacher and children/adolescent ratings

## Correlated Trait–Correlated Method Minus One Analysis of the Convergent and Discriminant Validity of the Conners 3 Short Forms

journals.sagepub.com/home/asm



The Conners Rating Scales (Conners, 1989, 1997, 2008) that have a history of more than 30 years are used extensively in clinical and research settings for screening and diagnosis of attention-deficit hyperactivity disorder (ADHD) and comorbid disorders in children and adolescents (Izzo, Donati, & Primi, 2018). The rating scales are currently in its third edition, and are called Conners 3 (C 3; Conners, 2008). C 3 has comparable versions for completion by parents, teachers, and self-rating by children/adolescents (Conners, 2008). Short versions, called Conners 3 Short [C 3 (S)], of these measures have also been developed from their longer counterparts. The C 3 (S) for completion by parents, teachers, and self are called Conners 3–Parent Short [C 3–P (S); 45 items], Conners 3–Teacher Short [C 3–T (S); 41 items], and Conners 3–Self-Rating Short [C 3–SR (S); 41 items]. Both the C 3–P (S) and C 3–T (S) are used to rate children between 6 and 18 years of age. The C 3–SR (S) is completed by children/adolescents between 8 and 18 years. As the different respondent versions have been offered as parallel measures capable of providing similar

cross-informant scores from parents, teachers, and children/adolescents, an important psychometric property is their interrater agreement. For the C 3 (S) versions, the present study used a multiple traits by multiple methods (MTMM) approach, called the correlated trait–correlated method minus one [CT-C(M – 1); Eid, 2000; Eid, Lischetzke, Nussbeck, & Trierweiler, 2003] to examine interrater agreement. More specifically, based on ratings from mothers, teachers, and children/adolescents themselves, the study examined the extent of agreement (convergence) of the same traits across mother ratings of the C 3–P (S), teacher ratings of the C 3–T (S), and adolescents/children self-ratings on the C 3–SR (S) (referred to as convergent validity), and the extent to which the different traits in the C 3 (S) are distinct (referred to as discriminant validity).

All the C 3 (S) respondent versions have a mixture of content and validity scales. The same two validity scales are present in all three versions. There are positive impression (PI, six items; measuring overly positive responding) and

negative impression (NI, six items; measuring overly negative responding). The content scales of the C 3 (S) measure

<sup>1</sup>Cairnmillar Institute, Hawthorn East, Victoria, Australia

<sup>2</sup>Federation University Australia, Mount Helen, Victoria, Australia

<sup>3</sup>Royal Children's Hospital, Melbourne, Victoria, Australia

<sup>4</sup>The University of Melbourne, Melbourne, Victoria, Australia

<sup>5</sup>National and Kapodistrian University of Athens, Athens, Greece

#### Corresponding Author:

Rapson Gomez, Cairnmillar Institute, 391-393 Tooronga Road, Hawthorn East, Victoria 3123, Australia.

Email: rapson.gomez@cairnmillar.edu.au

psychopathology, learning, cognitive, and relationship problems. The C 3-P (S) has six content scales, measuring inattention (IN; five items), hyperactivity/impulsivity (HY; six items), learning problems (LP; five items), executive functioning (EF; five items), aggression (AG; five items), and peer relations (PR; five items). The C 3-T (S) has five content scales, measuring IN (five items), HY (six items), LP/EF (six items of which four are LP items), AG (five items), and PR (five items). The C 3-SR (S) has five content scales, measuring IN (six items), HY (five items), LP (five items), AG (six items), and family relationships (FR, five items). Thus, the different versions differ slightly in the content scales. Also, the same scales across the versions differ in item content, and many comparable items across versions are worded slightly differently.

The psychometric properties of the C 3 (S) are comprehensively summarized in the C 3 manual (Conners, 2008). Given this, the major properties, especially those that have direct relevance to the present article (factor structure, internal consistency reliability, influence of age, and gender on ratings of the C 3 (S) items/scales, and interrater agreement) are summarized here. According to the C 3 manual, confirmatory factor analysis (CFA) of the C 3 (S), based on the items forming the content scales found support for the six-, five-, and five-factor theorized oblique models for the C 3-P (S), C 3-T (S), and C 3-SR (S), respectively. The internal consistency (Cronbach's alpha) reliabilities of the scales in the different versions ranged from .85 to .92 for the scales in the C 3-P (S), .87 to .94 for the scales in the C 3-T (S), and from .77 to .89 for the scales in the C 3-SR (S). All these values are well above .70 that is generally considered the minimum level for acceptable internal consistency reliability (Nunnally, 1978).

For all three C 3 (S) versions, the scoring involves computing the total raw scores for the scales by adding the responses of the items in the respective scales. The raw scores are then converted to *T* scores. According to the C 3 manual, for the C 3-P (S) versions, the scores for IN, HY, and EF are affected by gender (higher for males), and the score for HY is affected by age (higher in younger children). For the C 3-T

(S), all scale scores are affected by gender (higher for males), and except for AG, age affected all scale scores (higher in males). For the C 3-SR (S), none of the scale scores are affected by gender, and except for IN and AG, age affected the other scale scores. Given these findings, the *T* scores obtained from the raw scores for all the C 3 (S) scales are based on the age and gender of an individual.

In terms of interrater agreement, the manual reports that the across-informant correlations for parent and teacher ratings for IN, HY, LP (EF/LP for teachers), AG, and PR were .66, .61, .66, .52, and .62, respectively. The across-informant correlations for parent and adolescent/children self-ratings for IN, HY, LP, and AG were .59, .50, .66, and .53, respectively; and the across-informant correlations for teacher and adolescent/children self-ratings for IN, HY, LP, and AG were .51, .44, .57, and .42, respectively. Based on these values, the authors concluded that there are moderate levels of consistencies (agreements) across different informants (Conners, 2008). It is worth noting that the moderate across-informant agreements for the C 3 (S) is consistent with findings involving other child and adolescent measures (Choudhury, Pimentel, & Kendall, 2003; Grills & Ollendick, 2002). The meta-analysis by Achenbach, McConaughy, and Howell (1987) that involved 119 studies found low to moderate agreement between different informants, including parents, children, and teachers for ratings of social, emotional, and behavior problems in children and adolescents.

In terms of clinical utility, although the C 3 (s) measures were developed for screening and diagnosis of ADHD and disorders commonly comorbid with ADHD, at present there is no data on how the different C 3 (S) scales are related to children and adolescents diagnosed with various psychological disorders, including ADHD. Thus, examination of this would be valuable. Another valuable psychometric information for the C 3 (S) is interrater agreement. This is because comparable scales across the C 3-P (S), C 3-T (S) and C 3-SR (S) have been offered as parallel measures capable of providing similar cross-informant scores from parents, teachers, and adolescents/children, respectively. It is argued here that although interrater agreements data for different pairs of respondents are provided in the C 3 manual, they are likely to be confounded. This is because the data provided in the manual are based on observed scores. Such data are problematic (Willard, Conklin, Huang, Zhang, & Kahalley, 2016). The problem is best understood within the context of classical test theory (Lord & Novick, 1968). Classical test theory suggests that a raw observed score comprises variance for three components: trait, method, and random measurement error. Trait variance is the variance for the construct being measured while method variance is the systematic variance specific to the method

used to collect the information on the construct. Random measurement error variance includes all other variance and is unrelated to method or trait variance. Both method and random measurement effects are considered problematic when evaluating convergence because these effects distort (either increase or decrease) the relations among the trait constructs (Campbell & Fiske, 1959). The major problem with observed scores is that they comprise trait, method, and error variances together. Given that the interrater agreement scores in the C 3 manual are all based on observed scores, it can be argued that existing across-informant findings for the C 3 (S) are confounded with both method and error variances.

CFA methods, generally referred to MTMM, are available that allow the separation of trait, method, and error variances (Lance, Noble, & Scullen, 2002). In general, the MTMM approach involves data for two or more traits measured by two or more methods (or informants). The original Campbell and Fiske's (1959) MTMM approach evaluates convergent and discriminant validity using observed scores and therefore does not control for method and error variance. When a CFA framework is used, the analysis uses latent scores and is therefore free of error variance (Lance et al., 2002). CFA-based MTMM models allow evaluation of the convergence of a measure (trait variance), after taking into account both error and the method variance (systematic variance that is specific to the method used to collect information), and thereby can be expected to provide a more accurate estimation of convergent and divergent validities.

Two commonly used MTMM models are the correlated trait–correlated method (CT-CM) and the correlated trait–correlated uniqueness (CT-CU) models. These models allow evaluation of the convergence of a measure after taking into account the method and error variance in it. In the CT-CM model, the same subscales from the different methods load on their own trait factor. Also, all the subscales from the same method load on their own method factor, with the different trait factors correlating with each other, and the different method factors correlating with each other. Trait and method factors are not correlated. In the CT-CU model, the trait factors are similar to that for CT-CM model. However, the method effect is modelled by within method correlated error variances. For both models, support for convergent validity is inferred if there are relatively high trait variances for the subscales. In the CT-CM model, high method effect is inferred if there are relatively high variances for the different subscales within methods. In the CT-CU model, high method effect is inferred if there are relatively high correlations for the different error variances within methods. In both models, low correlations between the latent trait factors are taken as

evidence for their discriminant validity (Lance et al., 2002).

Despite the wide scale used of the CT-CM and the CT-CU methods (Burns & Haynes, 2006), there are problems with these methods (Eid et al., 2008; Geiser, Eid, & Nussbeck, 2008; Höfling, Schermelleh-Engel, & Moosbrugger, 2009; Nussbeck, Eid, Geiser, Courvoisier, & Lischetzke, 2009). Eid et al. have suggested that a key issue that has to be considered when selecting a CFA model for MTMM analysis is the types of methods in the model. Methods can be either interchangeable (all respondents have the same access to the target, and therefore rate the target from the same perspective) or structurally different (all raters have different access to the target, they would respond from different perspectives). According to these terms, when the C 3 (S) is completed by children/adolescents, mothers, and teachers, the data constitute structurally different methods. According to measurement experts, the CT-CU is not appropriate for structurally different methods.

Although the CT-CM can be used, there are a number of substantive and psychometric problems with this approach (Eid, 2000; Eid et al., 2003). Among them are that if there is a general trait factor across the different traits for a method, then the variance for this trait component will be incorporated as part of the variance of the method factor, thereby confounding results; and its application often leads to inadmissible solutions.

According to Eid et al., the CT-C(M – 1) model, which is derived from the CT-CM model, is the preferred model for structurally different methods (Eid et al., 2008; Geiser et al., 2008; Nussbeck et al., 2009). Such models have been used to examine the convergent and discriminant validities for other child/adolescent psychopathology measures, such as the Disruptive Behavior Rating Scale (Gomez & Gomez, 2015), the Strengths and Difficulties Questionnaire (Gomez, 2014), and the Achenbach System of Empirically Based Assessment (Gomez, Vance, & Gomez, 2014).

In the CT-C(M – 1) model, one of the methods is selected as the reference method (Eid, 2000; Eid et al., 2003). The true scores (consistency coefficients) of the reference method indicators are used to predict the true scores (consistency coefficients) of indicators of the other methods or nonreference methods. The consistency coefficient of the reference method is its true-score variance, while the consistency coefficients of the nonreference methods are the amount of variance in them that are predicted by the true score of the reference method. For any scale, the consistency coefficients of the nonreference methods indicate the convergence of the nonreference methods with the reference methods. The proportion of the true-score variance in the nonreference methods that are not predicted by the true score

of the reference method are their method-specific variance or method-specific coefficients.

Given that the interrater agreement scores for the C 3 (S) in the C 3 manual are likely to be confounded with both method and error variance, the first aim of the current study was to examine the convergent and discriminant validity of the C 3 (S) for ratings provided by adolescent/children [on the C 3-SR (S)], mothers [on the C 3-M (S)], and teachers [on the C 3-T (S)]. We focused on only the content scales as they are the ones that are substantively meaningful from a clinical viewpoint. We first used the CT-CM model. However, as this model produced out of range value for one of its indicators (error variance been negative), this model was deemed inadmissible. Consequently, we used the CT-C(M - 1) model. A problem with the CT-C(M - 1) model is that it is not symmetrical as the meaning of the parameters of the model depends on the method chosen as the reference standard (Geiser et al., 2008). If we choose mother ratings as the reference method in the CT-C(M - 1) model, it means that we are evaluating the convergence of mother ratings with teacher ratings and self-ratings. This analysis will not show how teacher ratings and self-ratings converge with each other. For this, we need to conduct a second analysis in which teacher ratings is the reference method and self-ratings is one of the nonreference method, or self-ratings is the reference method and teacher ratings is one of the nonreference method. Either of these models will show how teacher ratings and self-ratings converge with each other. Thus, if we wish to know how mother, teacher, and self-ratings converge with each other, we need to, at the very least, conduct two CT-C(M - 1) models. In one model, mother ratings could to be used as the reference method, and teacher and self-ratings as the nonreference methods, and in the other model, teacher ratings (or alternatively self-ratings) is the reference method and the remaining two methods are the nonreference methods. Given this, we conducted two different CT-C(M - 1) analyses. In the first analysis (MTMM Analysis 1), mother rating was used as the reference method. In the second analysis (MTMM Analysis 2), teacher rating was used as the reference method. Therefore, conducting these complementary analyses enabled examination of the inter convergence of all three pairs of respondents: mother ratings with teacher ratings (in MTMM Analyses 1 or 2), mother ratings with self-ratings (in MTMM Analysis 1), and teacher ratings with self-ratings (in MTMM Analysis 2). Based on exiting data, the expectation was to find at least moderate support for convergent validity across the different respondent pairs. Some support for discriminant validity of the C 3 (S) traits was expected. Discriminant validity for the different method factors was also expected. We also examined the external validity of the trait factors in the CT-C(M - 1) model by examining their correlations with anxiety, depressive, ADHD, and oppositional defiant

disorder/conduct (ODD/CD) diagnoses. The findings from such an evaluation would not only provide additional test of the construct validity of the C 3 (S) measures but also valuable information on utilization of the C 3 (S) measures for clinical diagnosis.

## Method

### Participants

The data for all participants were collected archivally from the Academic Child Psychiatry Unit (ACPU) of the Royal Children's Hospital, Melbourne, Australia. The ACPU is an out-patient psychiatric unit that provides services for children and adolescents with behavioral, emotional, and learning problems. Referrals are generally from other medical services, schools, and social and welfare organizations. All parents and children were informed that the clinic would provide diagnosis and appropriate treatment, and that assessment will be over 2 days, covering a range of tests involving the parents, children/adolescents, and teachers. They were informed that all data collected would be kept in an unidentifiable form in a secure database and (if consent was given) used to support future research.

Given the age range recommended for the application of the C 3-SR (S), for the current study, we used the records of children and adolescents aged between 8 and 18 years. The data comprised retrospective referrals between 2004 and 2017, who had been interviewed for clinical diagnosis. An individual was selected for inclusion in the study if that individual had ratings for the C 3 (S), completed by mothers, teachers, and self. Apart from this and the age criteria, no other inclusion/exclusion criterion was applied when selecting participants for the study. In all, there were 529 children and adolescents, comprising 70.5% males and 29.5% females. The overall mean age of participants was 11.75 years ( $SD = 2.97$  years).

Supplementary Table S1 (all supplemental materials are available in the online version of the article) shows the means and  $SD$  of the raw scores for all C 3 (S) scales used in the CT-C(M - 1) model. For all the C 3 (S) versions, the conversion of raw scores to  $T$  scores is based on age and gender. Since the mean age of the entire sample was 11.75 years and the mean scores for all the C 3 (S) scales in the parent, teacher, and self-report are higher for males than females, we used the normative scores for 12-year-old males to interpret the elevation of the raw scores. Based on this standard, all raw scores, except the LP scale of C 3-T (S) were either very elevated ( $T$ -score = 70+) or elevated ( $T$ -score = 60 to 69; Conners, 2008). The raw score for the teacher LP scale was average (40 to 59).



For the current study, all the clinical diagnoses for children and adolescents were derived from the Anxiety Disorders Interview Schedule for Children–Parent Version (ADISC-IV-P; Silverman & Albano, 1996) described below. The C 3 (S) measures were not used for facilitating clinical diagnosis. Supplementary Table S2 shows the percentages of different categories of disorders (any Anxiety Disorders, Dysthymia/Major Depressive Disorder, ADHD, and ODD/CD) for the participants. Any Anxiety Disorder includes Separation Anxiety, Social Phobia, Specific Phobia, Panic, Agoraphobia, Generalized Anxiety, Obsessive–Compulsive, and/or Posttraumatic Stress disorders. As shown in Table S2, there were high frequencies for all four categories of disorders. There were 26.8% of participants with all four disorder categories and 34.6% of participants had three disorder categories, with the highest comorbidities being for any anxiety disorder with ADHD and ODD/CD. For comorbidity, disorders were not distinguished as primary and secondary. There were 23.1% with two disorder categories, with the ADHD and ODD/CD being the highest comorbidity. Only 11.95% of participants had a single disorder, and 1.9% had no disorder.

In relation to socioeconomic related factors, the percentages of father employment status were as follows: employed = 75.1%, home duties = 2.5%, pensioner/retired = 5.7%, unemployed = 9.1%, others/unknown = 7.6%. The percentages of father highest education level were as follows: tertiary = 24.2%, high school/some years in secondary school or equivalent = 10.1%, technical certificate or equivalent = 24.6%, and primary school = 2.4%. The percentages of mother employment status were as follows: employed = 46.2%, home duties = 35.2%, pensioner/retired = 6.8%, unemployed = 3.4%, others/unknown = 6.4%. The percentages of mother highest education level were as follows: tertiary = 31.2%, high school/some years in secondary school or equivalent = 13.4%, technical certificate or equivalent = 24.61%, and primary school = 0.6%. Thus, most fathers and mothers of participants were employed, and more than two third of participants had fathers and mothers who had attended at least secondary school. In terms of parental relationship, about 50.1% were living together and 49.9% were separated or divorced. More than two thirds (62%) of participants were from families with income less than \$50,000 per year.

## Measures

*Anxiety Disorders Interview Schedule for Children–Parent Version* (Silverman & Albano, 1996). The ADISC-IV-P was

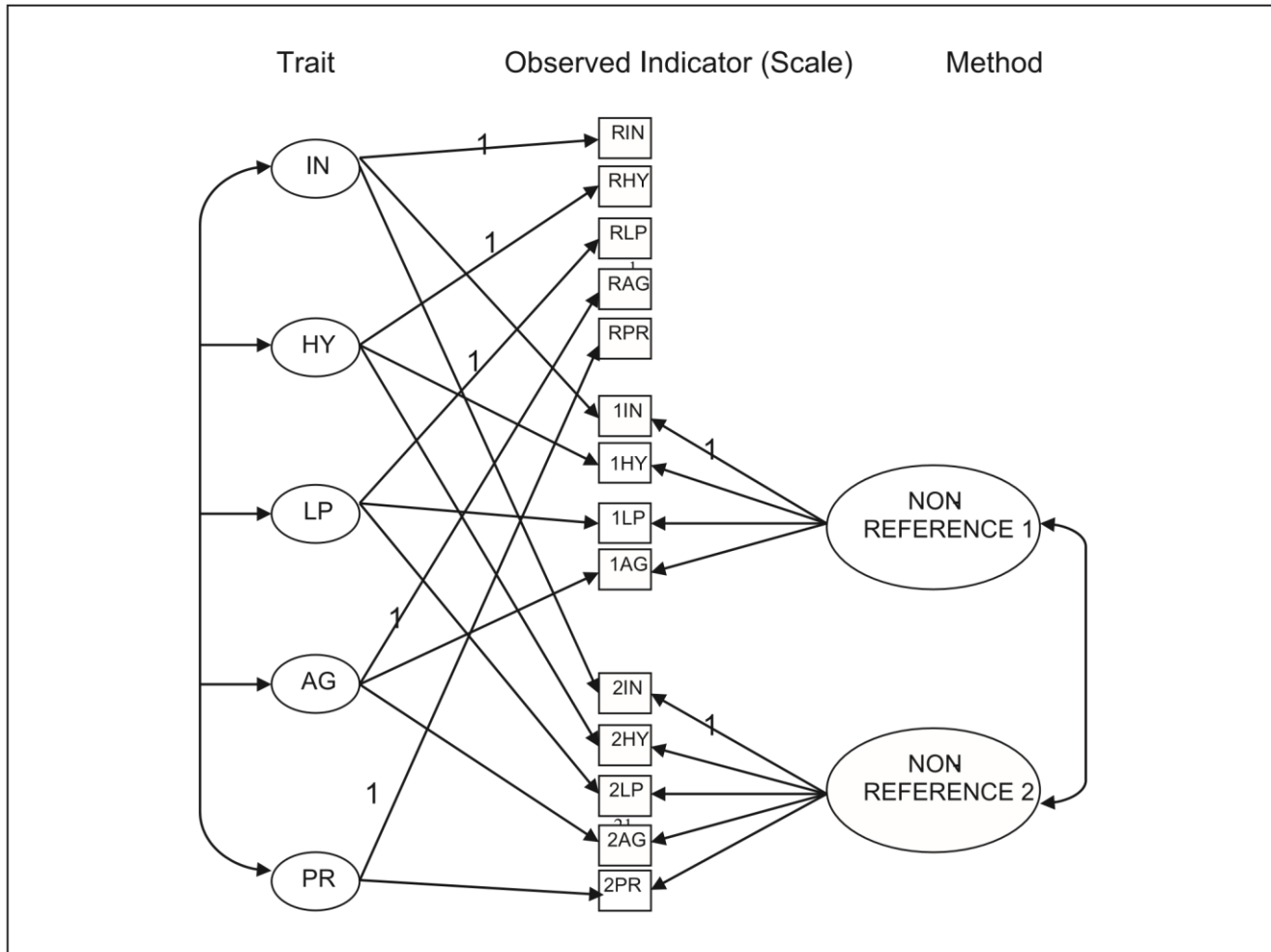
used for diagnosis, and these diagnoses were also used for examining the external validity of the trait factors in the CT-C(M – 1) model. The ADISC-IV-P is a semistructured interview, based on the *DSM-IV-TR* diagnostic system (American Psychiatric Association, 2000). It has been designed to facilitate the diagnosis of major childhood disorders. The ADISC-IV-P guidelines for diagnosis are that the child/adolescent be given a diagnosis of all disorders meeting the diagnostic criteria, and not in terms of primary and secondary disorders. Thus, all disorders that an individual qualified for were seen as equally applicable to that individual. The scores of ADISC-IV-P have sound psychometric properties (Silverman, Saavedra, & Pina, 2001). Test–retest reliability for the ADISC-IV-P scores over a 7- to 14-day interval has shown good to excellent reliability. Kappa values for interview with children between 7 and 16 years ranged from .61 to .80 (Silverman et al., 2001).

*Conners 3rd Edition Short [C 3 (S); Conners, 2008]*. As the parent, teacher, and self-report versions of the C 3 (S) was described comprehensively in the introduction, this section will only provide additional information not provided in the introduction. For all C 3 (S) versions, respondents indicate the degree or frequency of each behavior described in the item on a scale of 0 (not true at all), 1 (just a little true), or 2 (pretty much true), or 3 (very much true). The standard rating period is 1 month for all three versions. Scale scores are derived by summing the responses of the items in the respective scales, and then converting these raw scores to *T* scores, based on age and gender. In view of this, the raw scales scores were used in all the CT-C(M – 1) models. The internal consistency (Cronbach’s alpha) reliabilities of the scales in the different versions ranged from .82 to .92 for the scales in the C 3-P (S), .75 to .89 for the scales in the C 3-T (S), and from .71 to .93 for the scales in the C 3-SR (S). All these values are well above .70 that is generally considered the minimum level for acceptable internal consistency reliability (Nunnally, 1978).

As mentioned previously, for the CT-C(M – 1) analyses, we focused on only the content scales as they are the ones that are substantively meaningful from a clinical viewpoint. Also as already mentioned, the C 3-P (S) has content scales for IN, HY, LP, EF, AG, and PR. The C 3-T (S) has content scales IN, HY, LP/EF, AG, and PR; and the C 3-SR (S) has content scales for IN, HY, LP, AG, and FR. Given that the scales across the three versions differ in items content, and that the same items across the versions are worded differently, we conducted the CT-C(M – 1) analyses at the scale level. In this respect, as the CT-C(M – 1) model requires comparable scales across the different respondents, we used the LP/EF

teacher scale as a measure of teacher LP (given that four of its six items were LP items), and we did not include the FP scale

both parents and children gave permission for all relevant data collected by the ACPU of the RCH or provided by others to



**Figure 1.** Conceptual path diagram (assuming nonreference Method 1 to be self-ratings) of the correlated trait–correlated method minus one model, used in the study.

Note. IN = inattention; HY = hyperactivity/impulsivity; LP = learning problems; AG = aggression; PR = peer relations; R = reference method.

from the C 3-SR (S) in our CT-C(M – 1) models. Thus, our CT-C(M – 1) models had three methods (mother, teacher, and self-ratings) with five indicators for mother and teacher ratings (the total raw scale scores for IN, HY, LP for mother and EF/LP for teacher, AG, and PR) and four indicators for self-ratings (the total raw scale scores for IN, HY, LP, and AG). Such an asymmetry does not pose a problem for a CT-C(M – 1) model.

### Procedure

Children and parents participated in separate interviews and testing sessions with breaks over 2 days. Information was also obtained from teachers using various checklists and questionnaires. In all cases, parental and child consent forms were completed prior to the assessment. The consent from

be used in future research, and was approved by the RCH ethics committee as part of our group's comprehensive examination of psychopathology in children and adolescents. The data collected covered a comprehensive demographic, medical (primarily neurological and endocrinological), educational, psychological, familial, and social assessment of the child and his or her family. All psychological data were collected by research assistants, who were advanced doctoral students in clinical psychology, and under the supervision of two registered clinical psychologists.

The research assistants were provided with extensive supervised training and practice by the two psychologists prior to them collecting data. Training of the ADISC-IV-P included observations of it being administered by the psychologists. The research assistants commenced administering the ADISC-IV-P only after they attained

competence in its administration, as assessed by the two registered psychologists. There was adequate interrater reliability for the diagnoses made between the research assistants and the psychologists, and between research assistants (average kappa value across all diagnoses = .88).

Standard procedures were used for the administration of all measures. Where necessary, researchers read the C 3 (S) items to participants (approximately 5% of the sample). Approximately 85% of the parent ADISC-IV-P interviews involved mothers only, and the rest involved fathers only or both fathers and mothers together. Using the categorical data from the parent ADISC-IV-P, clinical diagnosis was determined by two consultant child and adolescent psychiatrists who independently reviewed the data. The interrater reliability for diagnoses of the two psychiatrists was high (kappa = .90).

### Statistical Procedures

Figure 1 shows a conceptual representation of the CT-C(M - 1) model tested in this study. In line with model specifications, all the indicators (IN, HY, LP, AG, and PR) of the reference method (mother in MTMM Analysis 1, and teacher in MTMM Analysis 2) were linked to their appropriate traits factors and not to any method factor. Also, these indicators for the nonreference methods (self-ratings and teacher in MTMM Analysis 1, and self-ratings and mother in MTMM Analysis 2) were linked to the appropriate traits factors and to their method factors. The trait factors correlated with each other, and the method factors correlated with each other. Methods and trait factors did not correlate. This CT-C(M - 1) models estimated 48 parameters. This means that with a sample size of 528, there were about 11 participants for every parameter estimated (ratio of 11:1).

In a CT-C(M - 1) model, the percentages of consistency coefficients and method-specific coefficients for the indicators can be examined for both observed and true scores (see Eid et al., 2003, Appendix A for appropriate formulas for computation of these). The square roots of the consistency coefficients or latent correlations of the indicators reflect the latent correlations between the true scores of reference and nonreference methods, with higher values indicating more convergence. Convergent validity of an indicator is inferred if it has large latent correlation and significant consistency coefficient. If in this instance, the indicator has a larger consistency coefficient than methodspecificity coefficient, then it means that there is good support for its convergent validity. For this study, support for convergence was inferred if the latent correlation was at least .70 between different respondent pairs for a trait since a correlation of .70 reflects approximately 50% of shared variance between the traits

in question. Additionally, it was also necessary for the consistency coefficient to be significant, and either relatively higher or comparable to the method-specificity coefficient. Although the meaning of a factor loading varies by research context, the degree of correlations of the trait factors indicate the discriminant validity of the traits (as reflected in the reference method), whereas the degree of correlations between the method factors indicate the discriminant validity of the methods. In both cases, low values support their discriminant validity. For this study, correlations of less than .70 were taken as support for the discriminant validity of the traits. Independent of the evaluation of convergent validity, we examined the level of convergence of the nonreference method with the reference method. Although the meaning of a factor loading varies by research context, Garson (2013) has suggested that factor loadings that are less than .4 (or consistency coefficient of .16) are “weak,” loadings from .4 to .6 (consistency coefficients from .16 to .36) are “moderate,” and loadings above .6 are “strong” (consistency coefficients above .36). These consistency coefficient values were used in the current study to quantify the level of convergence.

We examined the relations of the trait factors in the CT-C(M - 1) model with clinical disorders by extending the CT-C(M - 1) model shown in Figure 1 to include correlation paths between each of the trait factors to anxiety disorders, depressive disorders, ADHD, and ODD/CD. The strength of all correlations was interpreted using the guidelines proposed by Hemphill (2003) for correlation effect sizes:  $<.2$  = small,  $.2$  to  $.3$  = medium or moderate, and  $>.30$  = large.

All the CFA models in the study were analyzed using *Mplus* (Version 7) software (Muthén & Muthén, 2010). Robust maximum likelihood was used for the CT-C(M - 1) estimations. The robust scaled chi-square statistic (called Satorra-Bentler or  $S-B\chi^2$ ), the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA) were used to ascertain model fit. The guidelines suggested by Hu and Bentler (1998) are that CFI values close to .95 or above are taken as indicating good model-data fit, and values of .90 and  $<.95$  are taken as acceptable fit. For the RMSEA, values close to .06 or below can be taken as good fit and values close to .07 to  $<.08$  as moderate fit.

## Results

### Missing Data

Out of a total of 7,292 scores in the CT-C(M - 1) model (14 scales  $\times$  528 participants), there were 208 scores missing (i.e.,



around 2.85%). Maximum likelihood (direct ML) was used to handle missing data (Brown, 2006).

### *Preliminary Analyses*

Prior to the MTMM analysis the originally proposed factor structure of the C 3-P (S), C 3-T (S), and C 3-SR (S) were examined. As the C 3-P (S) and the C 3-SR (S) did not include the EF and FR scales, respectively, in the CT-C(M - 1) models, we examined support for a five-factor model (without the original EF scale) for the C 3-P (S), and also a four-factor model (without the original FR scale) for the C 3-SR (S). The fit indices for the modified five-factor model for the C 3-P (S) were  $WLSMV\chi^2$  (degrees of freedom [ $df$ ] = 289) = 765.24,  $p < .001$ , CFI = .974, and RMSEA = .057, 90% confidence interval (CI) [.052, .062]. The fit indices for the proposed five-factor model for the C 3-T (S) were  $WLSMV\chi^2$  ( $df$  = 289) = 845.90,  $p < .001$ , CFI = .967, and RMSEA = .074, 90% CI [.069, .080]. The fit indices for the modified four-factor model for the C 3-SR(S) were  $WLSMV\chi^2$  ( $df$  = 203) = 573.72,  $p < .001$ , CFI = .959, and RMSEA = .066, 90% CI [.059, .072]. Taken together, these findings can be interpreted as indicated good fit for the factor models in the various C 3 (S) measures that correspond to the scales from these measures used in the CT-C(M - 1) models in the current study.

### *MTMM Analysis 1 (Mother as the Reference Method)*

The fit indices for the CT-C(M - 1) model with mother as the reference method (MTMM Analysis 1) were  $S-B\chi^2$  ( $df$  = 57) = 149.91,  $p < .001$ , CFI = .962, and RMSEA = .056, 90% CI [0.045, 0.066]. Thus, both the CFI and RSMEA values indicated good model-data fit.

Table 1 provides the variance components of the traits (IN, HY, LP, AG, and PR) and method (mother, teacher, and self) factors, and latent correlations of the IN, HY, LP, AG, and PR indicators. As shown, with the exceptions of the teacher and self-ratings for the IN and HY indicators, the

**Table 1.** Variance Components in the CT-C(M – 1) Model With Mother as the Reference Method.

Ratings	Observed variables			Trait-score variables		
	Reliability	Consistency	Method specificity	Consistency	Method specificity	Latent correlation
<b>Inattention</b>						
Mother	.78	.78***	.00	1.00	.00	1.00
Teacher	.84	.20***	.64***	.24	.76	.49
Self	.84	.24***	.61***	.28	.72	.53
<b>Hyperactivity/impulsivity</b>						
Mother	.76	.77***	.00	1.00	.00	1.00
Teacher	.73	.28***	.45***	.38	.62	.62
Self	.73	.17***	.56***	.23	.77	.48
<b>Learning problems</b>						
Mother	.99	.99***	.00	1.00	.00	1.00
Teacher	.57	.30***	.27***	.53	.47	.72
Self	.62	.39***	.22***	.64	.36	.80
<b>Aggression</b>						
Mother	.66	.66***	.00	1.00	.00	1.00
Teacher	.64	.31***	.32***	.49	.51	.70
Self	.49	.31***	.18***	.62	.38	.78
<b>Peer relations</b>						
Mother	.71	.71***	.00	1.00	.00	1.00
Teacher	.57	.38***	.19***	.66	.34	.81

Note. CT-C(M – 1) = correlated trait–correlated method minus one. \*\*\* $p < .001$ .

**Table 2.** Correlations of the Trait and Method Factors in the CT-C(M – 1) Model With Mother as the Reference Method.

	1	2	3	4	5	6	7
<b>Trait</b>							
Inattention (1)	—						
Hyperactivity/impulsivity (2)	.82*	—					
Learning problems (3)	.69*	.41*	—				
Aggression (4)	.49*	.66*	.23*	—			
Peer relations (5)	.46*	.45*	.35*	.55*	—		
<b>Method</b>							
Self (6)						—	
Teacher (7)						.27*	—

Note. CT-C(M – 1) = correlated trait–correlated method minus one. \* $p < .001$ .

latent correlations for all the other indicators for teacher and self-ratings (LP, AG, and PR) were all at or above .70. The latent correlations for teacher and self-ratings for the IN and HY indicators were less than .70. The consistency coefficients for all indicators were significant. Additionally, the consistency coefficients for teacher and self-ratings of LP,

AG, and PR were either close to or relatively more than their method-specificity coefficients, and the consistency coefficients for teacher and self-ratings of IN and HY were relatively low compared with their method-specificity coefficients. These findings can be interpreted as showing support for the convergence of teacher and self-ratings of LP, AG, and PR with mother ratings of these traits. Also, there is no support for teacher and self-ratings of IN and HY with mother ratings of these traits. Based on the guidelines suggest by Garson (2013, i.e., consistency coefficient of  $<.16$ ) are “weak,” consistency coefficients from .16 to .36 are “moderate,” and consistency coefficients above .36 are strong), the magnitudes of the consistency coefficients were moderate for IN, HY, and AG for both teacher ratings and self-ratings. For LP, the magnitude of the consistency coefficient for teacher ratings was moderate, whereas it was strong for self-ratings. The magnitude of the consistency coefficient for teacher ratings of PR was moderate.

Table 2 shows the correlations between the trait factors and the method factors. As shown, with the exception of the correlations between IN and HY, all other correlations were less than .70 (the criteria for adequate discrimination). Thus, the findings indicated support for the discriminant validity of the C 3 (S) factors, except IN and HY. The correlation between self-ratings and teacher method factors was .27. As this value is less than .70, the discriminant validity of these method factors can be inferred.

**MTMM Analysis 2 (Teacher as the Reference Method)**

The fit indices for the CT-C(M – 1) model with teacher as the reference method (MTMM Analysis 2) were  $S-B\chi^2 (df = 57) = 213.45, p < .001, CFI = .936,$  and  $RMSEA = .072,$  90%

CI [.062, .083]. Thus, both the CFI and RSMEA values indicated adequate model-data fit.

Table 3 provides the variance components of the traits (IN, HY, LP, AG, and PR) and method (mother, teacher, and self) factors, and latent correlations of the IN, HY, LP, AG, and PR indicators. As shown, with the

**Table 3.** Variance Components in the CT-C(M – 1) Model With Teacher as the Reference Method.

Ratings	Observed variables			Trait-score variables		
	Reliability	Consistency	Method specificity	Consistency	Method specificity	Latent correlation
<b>Inattention</b>						
Teacher	.81	.81	.00	1.00	.00	1.00
Self	.82	.25***	.57***	.31	.69	.56
Mother	.77	.29***	.48***	.38	.62	.62
<b>Hyperactivity/impulsivity</b>						
Teacher	.82	.82***	.00	1.00	.00	1.00
Self	.77	.16***	.61***	.20	.80	.45
Mother	.68	.29***	.39***	.43	.57	.66
<b>Learning problems</b>						
Teacher	.61	.61***	.00	1.00	.00	1.00
Self	.67	.44***	.23***	.66	.34	.81
Mother	.84	.53***	.30***	.64	.36	.80
<b>Aggression</b>						
Teacher	.80	.80***	.00	1.00	.00	1.00
Self	.49	.30***	.19***	.62	.38	.79
Mother	.50	.29***	.21***	.58	.42	.76
<b>Peer relations</b>						
Teacher	.92	.92***	.00	1.00	.00	1.00
Mother	.47	.31***	.15***	.67	.33	.82

Note. CT-C(M – 1) = correlated trait–correlated method minus one. \*\*\* $p < .001$ .

**Table 4.** Correlations of the Trait and Method Factors in the CT-C(M – 1) Model With Teacher as the Reference Method.

	1	2	3	4	5	6	7
<b>Trait</b>							
Inattention (1)	—						
Hyperactivity/impulsivity (2)	.85*	—					
Learning problems (3)	.66*	.32*	—				
Aggression (4)	.57*	.68*	.20*	—			
Peer relations (5)	.46*	.43*	.31*	.62*	—		
<b>Method</b>							
Self (6)						—	
Parent (7)						.31*	—

Note. CT-C(M – 1) = correlated trait–correlated method minus one. \* $p < .001$ .

exceptions of the mother and self-ratings for the IN and HY indicators, the latent correlations for all the other indicators for mother and self-ratings (LP, AG, and PR) were all at or above .70. The latent correlations for mother and self-ratings for the

IN and HY indicators were less than .70. The consistency coefficients for all indicators were significant. Additionally, the consistency coefficients for mother and self-ratings of LP, AG, and PR were either close to or relative more than their method-specificity coefficients, and the consistency coefficients for mother and self-ratings of IN and HY were relatively low compared with their method-specificity coefficients. These findings can be interpreted as showing support for the convergence of mother and self-ratings of LP, AG, and PR with teacher ratings of these traits. Also, there was no support for mother and self-ratings of IN and HY with teacher ratings of these traits. The magnitudes of the consistency coefficients for IN and AG were moderated for both mother ratings and self-ratings. For HY, it was weak for self-ratings and moderate for mother ratings. For LP, the

magnitude of the consistency coefficients for both mother and self-ratings were strong. The magnitude of the consistency coefficient for mother ratings of PR was moderate.

Table 4 shows the correlations between the trait factors and the method factors. As shown, with the exception of the correlations between IN and HY, all other correlations were less than .70 (the criteria for adequate discrimination). Thus, the findings indicated support for the discriminant validity of the C 3 (S) factors, except IN and HY. The correlation between self-ratings and mother method factors was .31. As this value is less than .70, the discriminant validity of these method factors can be inferred.

**Table 5.** Correlations of the Latent Trait and Method Factors Scores in the Mother Reference Method CT-C(M – 1) Model with Major Categories of Clinical Disorders.

Disorders	Trait factors					Method factors	
	IN	HY	LP	AG	PR	Teacher	Self
ADHD	.53***	.45***	.32***	.24***	.21***	.11	-.05
ODD/CD	.05	.20***	-.01	.52***	.18***	.14	.06
Depressive disorders	.05	-.05	-.01	.12*	.04	-.10	.07
Anxiety disorders	-.11*	-.03	-.05	-.01	.13**	.01	.09

Note. CT-C(M – 1) = correlated trait–correlated method minus one; ADHD = attention-deficit hyperactivity disorder; ODD/CD = oppositional defiant disorder/conduct; IN = inattention, HY = hyperactivity/impulsivity, LP = learning problems, AG = aggression, PR = peer relations.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

### Correlations of the Latent Trait and Method Factors with Different Disorder Categories

As the findings for the correlations of the latent trait factors with the different disorder categories were the same for the analysis with mother as the reference factor and teacher as the reference factor, we present here the findings for only the analysis with mother as the reference factor in the CT-C(M – 1) model. These correlations are presented in Table 5. As shown, all the C 3 (S) trait factors were significantly correlated with ADHD. The correlations involving IN, HY, and LP were of large effect sizes, and the correlations for AG and PR were of medium effect sizes (based on guidelines proposed by Hemphill [2003] for correlation effect sizes:  $<.2$  = small,  $.2$  to  $.3$  = medium or moderate, and  $>.30$  = large). ODD/CD correlated significantly with large effect size with AG, and small effect sizes with HY and PR. Depression disorders correlated with small effect size with AG, and any anxiety disorders correlated negatively with small effect size with IN, and positively with small effect size with PR. Taken together, the differential associations of the C 3 (S) traits with the different disorder categories provide good support for the external validity of the trait factors in the postulated CT-C(M – 1) model. In addition, the relatively large associations for ADHD with IN and HY, and ODD/CD with AG indicate that these scales are particularly useful for identifying ADHD and OD/CD, respectively. Table 5 also includes the correlations

of the teacher-ratings and self-ratings method factors with the different disorder categories. As shown in table, both the teacher ratings and self-ratings method factors were not significantly associated with any disorder category. For the CT-C(M – 1) model having teacher ratings as the reference method, the correlations for mother ratings with ADHD, ODD/CD, any depressive disorder, and any anxiety disorder were .11, .09, .08, and  $-.05$ , respectively. The correlations for self-ratings with ADHD, ODD/CD, any depressive disorder, and any anxiety disorder were .03, .06, .10, and .07, respectively. All the correlations for both mother ratings and teacher ratings were not significant.

### Discussion

For the CT-C(M – 1) analyses, we used the single method/trait indicator version of this model, with scale scores as indicators. As the scales across the three versions are not identical, the model for each version used in the CT-C(M – 1) analyses was not the theorized factor structures proposed for these versions. For all versions, we included the IN, HY, LP (LP/EF in the case of teacher ratings), and AG scales. Additionally, the PR scale was included for only mother and teacher rating, as there is no such scale for adolescent/ children self-ratings. Thus, in the CT-C(M – 1) model, the C 3-P and C 3-T measures comprised five scales [IN, HY, LP (more specifically LP/EF in the case of teacher rating), AG, and PR]; and the C 3-SR measure comprised four scales (IN, HY, LP, and AG). For the CT-C(M – 1) analyses, the findings showed that there was support for the convergence of teacher ratings of LP, AG, and PR with mother ratings of these traits, and self-ratings of LP and AG with mother ratings of these traits. Additionally, there was support for the convergence of self-ratings of LP and AG with teacher ratings of these traits. However, for IN and HY, there was no support for convergence for teacher and self-ratings with mother ratings, and self-ratings with teacher ratings. The findings also indicated support for discrimination between all trait factors, except IN and HY, and for self-ratings and teacher method factors, and mother and self-ratings method factors.

Independent of support or otherwise for convergence for the different traits across the different pairs of respondents, the findings indicated moderate levels of agreement for teacher and self-ratings with mother ratings for IN, HY, and AG, teacher and mother ratings for LP, teacher and self-ratings for IN and AG. These findings are consistent with the findings for these scales across these pairs of respondents as reported in the C 3 manual. The findings in the current study showed that the agreement for mother ratings with self-ratings for LP, and mother ratings with teacher ratings for PR were both strong. These findings are not consistent with the moderate agreement for these scales reported in the C 3 manual across these pairs of respondents. It is interesting to note that the moderate level of agreements across the different respondent pairs for most of the traits found in the current study is consistent with findings involving other child and adolescent measures (Achenbach et al., 1987; Choudhury et al., 2003; Grills & Ollendick, 2002), including studies using the CT-C(M - 1) approach (Gomez, 2014; Gomez et al., 2014).

Although most (but not all) of our interrater agreement findings are in general agreement with existing data provided in the C 3 manual, it is argued here that the findings in the current study are likely to be more accurate since they are based on a CFA MTMM approach. The findings reported in the manual are based on correlations involving observed scores. Such score included not only trait variance but also method and error variances. The inclusions of method and error variances confound convergence analysis because these effects distort (either increase or decrease) the relations among the trait constructs (Campbell & Fiske, 1959). CFA scores are latent scores that are free of error variances. Additionally, the scores in a CFA MTMM model (as used in the current study) partials out method variance also in the scores, thereby analyzing only trait variance. Since our convergence findings were based on only trait variance, they can be expected to be more accurate and reliable. Additionally, as the current study used the CT-C(M-1), it applied a CFA MTMM methodology that is well suited for structurally different methods as reflected in C 3 (S) ratings provided by mother, teacher and adolescents/children ratings.

To date, a number of explanations have been proposed for only moderate cross-informant agreement for child and adolescent measures. One explanation for cross-informant differences have been the observability of the behaviors in question, with higher informant agreement being proposed for more observable behaviors (De Los Reyes & Kazdin, 2005). IN and HY are highly observable behaviors. Thus, the lack of convergence for IN and HY across the different pairs of respondents in the current study is not consistent with this view. This raises the

possibility that the observability of the behavior may not be a major factor influencing the level of cross-informant agreement. Other explanations proposed for low to moderate crossinformant agreement relates to either real differences in children's and adolescents' behaviors at home, school, and other settings (situation specificity hypothesis), or differences in respondents' perceptions of children's and adolescents behaviors (bias hypothesis). Although the relative merits of these hypotheses could not be tested in the current study, existing data have provided more support for the situation specificity hypothesis (Achenbach et al., 1987).

The findings in this study have clinical implications and also implications for the use of the C 3 measures. First, in the current study, it can be assumed that adolescents/ children self-ratings of the C 3 (S) would most likely be cross-situational since their ratings cover all their different settings. If so, the findings indicate substantial more convergence for mother-adolescent than teacher-adolescent ratings of LP suggests that relative to teachers, mothers are more able to provide a broader judgment of adolescents and children's learning problems, and also that they are more able to identify adolescents and children's own perceptions of their learning problems. Second, since the amount of convergence for most (LP may be an exception) of the corresponding C 3 (S) scales across the different respondents was only moderate, it follows that the different respondents are not providing identical measures, and that they each have different information to offer. This highlights the need for researchers and clinicians to be cautious when interpreting scores derived from a single source, and conversely the need for obtaining C 3 (S) ratings from multiple sources. Indeed, Goodman (2001) showed that combining the information from parent, teacher, and adolescent self-rating resulted in the highest sensitivity in predicting psychiatric disorders, compared with using information from single sources or combinations of two sources. Third, as this study found no support for the discriminant validity between IN and HY, it can be argued that these scales are confounded with each other. Based on recent studies supporting the general factor in the ADHD bifactor model (Arias, Ponce, Martínez-Molina, Arias, & Núñez, 2016), one possibility is that together, IN and HY scales may constitute as an overall measure for ADHD rather than separate measures for IN and HY. However, this interpretation needs to be viewed cautiously as there are many conceptual and technical problems with the application and the interpretation of findings involving the bifactor model (Arias et al., 2016; Eid, Geiser, Koch, & Heene, 2017; Murray, & Johnson, 2013). Of particular relevance to our interpretation is that if there is an item that cross-loads, say on two latent factor, then the failure



to model this cross-loading, as is a requirement in a bifactor model, will result in artificially inflating the correlation between the two latent factors, as has been clearly demonstrated for the ADHD symptoms (Arias et al., 2016). This finding diminishes the argument that the IN and HY scales may constitute an overall measure for ADHD. Regardless, the lack of support for the discriminant validity between IN and HY found in this study underscores the need to be cautious when interpreting and integrating the IN and HY scores. Fourth, related to the lack of convergence for IN and HY, the study findings have implications for combining information on ADHD related symptoms provided by mothers, children/ adolescents, and teachers. They indicate the importance of obtaining and integrating information from multiple sources as this is likely to result in more informed diagnostic decisions than information based on either one or two of these respondents. However, there is no easy way to merge the information from them as their ratings for both IN and HY includes high levels of method variance. The implication here is that clinicians need to carefully and methodologically evaluate the degree to which reports of ADHD symptoms is influenced by method factors. For example, respondents would be asked appropriate questions for the different symptoms, and with such information, the clinician could explore common issues influencing the responses for IN and HY by different respondents, and also the responses for IN and HY across different respondents. Following this, they could evaluate the degree to which any reported symptom is indeed reflecting primarily the IN or HY symptoms and not method influences. In this respect, it is to be noted that a problem with MTMM models is how to interpret method factors. Although by convention, these factors are referred to as “method factors,” they should not necessarily be interpreted as such, and they may sometimes have a substantive interpretation for the measure (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), or simply be modeling the absence of strong invariance between the respondents’ ratings. However, in this respect, as our findings did not show any association for the mother, teacher, and self-ratings method factors with clinical diagnoses of ADHD, ODD/CD, depressive disorders, and anxiety disorders, we could rule out that our method factors are confounded with variances relevant for clinical diagnosis. Fifth, the findings here suggest that the LP and AG scales provide more consistent information across mother, teacher, and adolescent ratings than the other scales. In this respect, as for LP scale used for teacher ratings was the LP/ EF, it follows that it is the LP/EF teacher scale that is strongly convergent with mother and self LP scales. Sixth, the findings in the study also supported the external validity

of the latent trait factors in the CT-C(M – 1) model, with relatively stronger associations for IN and HY with ADHD, and AG with ODD/CD. This means that when the C 3-P (S), C 3-T (S), and C3-SR (S) are considered together, scores for IN and HY are useful in diagnosis of ADHD; and scores for AG are useful in diagnosis of ODD and CD.

In concluding, it need to be noted that the findings and interpretations made in the study need to be viewed with some limitations in mind. First, as already noted earlier, the scales used in the CT-C(M – 1) model were not based on the theorized factor structures of the different C 3 (S) versions. For example, the LP for teacher ratings was the LP/EF scale as there is no “pure” LP scale for this version. Additionally, PR was included for mother and teacher ratings only, as there is no such scale for self-ratings. Thus, the CT-C(M – 1) model for the different respondent versions do not correspond directly with the C 3 (S) proposed models. Second, the findings reported here are based on a single study. As a consequence, there is a need for crossvalidation of the findings before the findings can be generalized. Third, all the participants in this study were from the same clinic. Thus, it is possible that this may constitute an additional bias for the sample examined, limiting the findings and conclusions made in this study. Fourth, as this study was on clinic-referred children and adolescents, the applicability of the findings for children and adolescents in the general community cannot be assumed. Fifth, as the sample examined was highly heterogeneous and comorbid for a range of disorders, these may have confounded findings. Sixth, in the CT-C(M – 1) model the method selected as the reference has potential impact on the actual findings. It is important to keep this in mind when considering the findings. Although we reported CT-C(M – 1) analyses with mother and teacher as reference methods, it was not possible to do this with self-rating as the reference method since this measure did not have the PR scale. Thus, the findings here can be seen as specific to the C 3-P (S) and the C 3-T (S) as reference methods. Seventh, although the C 3-P (S), C 3-T (S), and C 3-SR (S) provide scale scores for IN, HY, LP, and AG, they do not have identical items. It is possible that this difference could have confounded the findings. It will be useful for future studies to examine samples from several clinics and from the general community in the same study, taking into all or some of the methodological limitations highlighted in this section.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Vasileios Stavropoulos  <https://orcid.org/0000-0001-6964-4662>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Achenbach, T. H., McConaughy, S. H., & Howell, C. Y. (1987). Child/adolescent behavior and emotional problems: Implications for cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Arias, V. B., Ponce, F. P., Martínez-Molina, A., Arias, B., & Núñez, D. (2016). General and specific attention-deficit/hyperactivity disorder factors of children 4 to 6 years of age: An exploratory structural equation modeling approach to assessing symptom multidimensionality. *Journal of Abnormal Psychology, 125*, 125-137.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- 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* (pp. 401-418). Washington, DC: American Psychological Association.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 2*, 81-105.
- Choudhury, M. S., Pimentel, S. S., & Kendall, P. C. (2003). Childhood anxiety disorders: Parent-child (dis)agreement using a structured interview for the *DSM-IV*. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 957-964.
- Conners, C. K. (1989). *Manual for Conners' rating scales*. North Tonawanda, NY: Multi-Health Systems.
- Conners, C. K. (1997). *Conners' rating scales-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Conners, C. K. (2008). *Conners rating scales* (3rd ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*, 483-509.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65*, 241-261.
- Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*, 541-562.
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods, 8*, 38-60.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods, 13*, 230-253.
- Garson, G. D. (2013). *Factor analysis*. Statistical Associates Publishing.
- Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods, 13*, 49-57.
- Gomez, R. (2014). Correlated trait-correlated method minus one analysis of the convergent and discriminant validities of the Strengths and Difficulties Questionnaire. *Assessment, 21*, 372-382.
- Gomez, R., & Gomez, A. (2015). Agreement of adolescent ratings with mother ratings and teacher ratings of ADHD symptom groups: A correlated trait-correlated method minus one analysis. *Personality and Individual Differences, 82*, 131-135.
- Gomez, R., Vance, A., & Gomez, R. M. (2014). Analysis of the convergent and discriminant validity of the CBCL, TRF and YSR in a clinic-referred sample. *Journal of Abnormal Child Psychology, 42*, 1413-1425.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 1337-1345.
- Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review, 5*, 57-83.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist, 58*, 78-79.
- Höfling, V., Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three approaches. *Methodology, 5*, 99-111.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
- Izzo, V. A., Donati, M. A., & Primi, C. (2018). Conners 3-SelfReport scale: An empirical support to the dimensionality of the content scales. *Clinical Child Psychology and Psychiatry*. Advanced online publication. doi:10.1177/1359104518757289
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228-244.
- Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 41*, 407-422.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nunnally, J. (1978). *Psychometric theory*. New York, NY: McGrawHill.
- Nussbeck, F. W., Eid, M., Geiser, C., Courvoisier, D. S., & Lischetzke, T. (2009). A CTC(M-1) model for different types of raters. *Methodology, 5*, 88-98.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A

Gomez et al.

critical review of the literature and recommended remedies.

*Journal of Applied Psychology*, 88, 879-903.

Silverman, W. K., & Albano, A. M. (1996). *Manual for the ADIS-IV C/P*. New York, NY: Psychological Corporation.

Silverman, W. K., Saavedra, L. M., & Pina, A. A. (2001). Test-retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for *DSM-IV*: Child and parent versions. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 937-944.

Willard, V. W., Conklin, H. M., Huang, L., Zhang, H., & Kahalley, L. S. (2016). Concordance of parent-, teacher- and self-report ratings on the Conners 3 in adolescent survivors of cancer. *Psychological Assessment*, 28, 1110-1118.



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Gomez, R; Vance, A; Stavropoulos, V

Title:

Correlated Trait-Correlated Method Minus One Analysis of the Convergent and Discriminant Validity of the Conners 3 Short Forms

Date:

2018

Citation:

Gomez, R., Vance, A. & Stavropoulos, V. (2018). Correlated Trait-Correlated Method Minus One Analysis of the Convergent and Discriminant Validity of the Conners 3 Short Forms. *Assessment*, 27 (7), pp.1463-1475. <https://doi.org/10.1177/1073191118803714>.

Persistent Link:

<http://hdl.handle.net/11343/227741>

---

File Description:

Accepted version