Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers

Lourdes Ezpeleta, Ph.D.[1,2]

Eva Penelo, Ph.D.[3]

[1]Unitat d'Epidemiologia i de Diagnòstic en Psicopatologia del Desenvolupament

[2]Departament de Psicologia Clínica i de la Salut

[3]Departament de Psicobiologia i Metodologia de les Ciències de la Salut

Universitat Autònoma de Barcelona (Spain)

Acknowledgements

Funding was from the Spanish Ministry of Science and Innovation grant PSI2009-07542 and grant PSI2012-32695 from the Ministry of Economy and Competitiveness. We should like to thank the participating schools and families.

Post-print version

Ezpeleta, L. & Penelo, E. (2015). Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers. European Journal of Psychological Assessment, 31, 45-53. doi: 10.1027/1015-5759/a000205

Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers

Abstract

Measurement invariance (metric/scalar) of oppositional defiant disorder (ODD) dimensions (negative affect, oppositional behaviour, and antagonistic behaviour) across sex and informants is tested. Parents and teachers of 622 preschool children from the general population answered a dimensional measure of ODD. ODD dimensions function similarly in boys and girls. Some differences were found by informant, indicating that the equivalence of the ratings of parents and teachers is not complete and that given the same underlying level of the latent trait, some parents' item scores were higher than those of teachers. Metric invariance was complete but scalar invariance was not attained. The results contribute evidence on the conceptualization of ODD as a source-specific disorder. The simultaneous use of ODD dimensions reported by parents and teachers must be considered in the context of a lack of complete measurement invariance, which implies that comparisons of observed means from parents and teachers are not readily interpretable.

KEYWORDS: dimensions; invariance; oppositional defiant disorder; preschool.

Oppositional defiant disorder is a highly prevalent disorder (Bufferd, Dougherty, Carlson, & Klein, 2011; Lavigne, Lebailly, Hopkins, Gouze, & Binns, 2009) that is accompanied by multiple comorbid disorders (Burke & Loeber, 2010; Lavigne, et al., 2001). Currently, ODD is considered a heterogeneous disorder that affects not only behaviour but also emotional dysregulation. In order to understand the nature of the varied comorbidity of the disorder, several theoretical or empirical dimensions have been proposed in samples of the general population from ages 3 to 16 (Rowe, Costello, Angold, Copeland, & Maughan, 2010; Stringaris & Goodman, 2009b). Specifically, Burke, Hipwell, and Loeber (2010) obtained three factors: 1) *negative affect*, containing the symptoms touchy, angry and spiteful; 2) *oppositional behaviour,* including loses temper, defies and argues, and 3) *antagonistic behaviour*, including annoys and blames. This structure was confirmed in a sample of clinically-referred boys and in a community sample of 5 to 8-year-old girls using dimensional measures of psychopathology. Recently, Ezpeleta, Granero, Osa, Penelo, and Doménech (2012) also confirmed these dimensions in parents' and teachers' ratings of 3-year-old preschoolers. In this sample, Burke's model showed better fit than alternative models (Rowe, et al., 2010; Stringaris & Goodman, 2009) which were also examined, the present study therefore focuses on this model. The identification of the factor structure of ODD symptoms, distinguishing several components of ODD, could prove highly advantageous in clinical contexts by helping to improve the understanding and prevention of ODD comorbidity.

These ODD dimensions have proved useful in the differential prediction of problems, and have shown predictive validity: *negative affect* is associated, both cross-sectionally and longitudinally, with emotional disorders, *oppositional behaviour* is related to disruptive behaviour disorders, and *antagonistic* behaviour is related to disruptive and mood disorders (Burke, et al., 2010; Ezpeleta, et al., 2012).

There is no information available about how these ODD dimensions function cross-informant (parents and teachers) or even cross-sex, and whether comparability of the ODD means in the different groups of responses is guaranteed. Measurement invariance deals with whether or not, under different conditions, measurements yield measures of the same attributes. Thus, this technique allows researchers to examine the equivalence of ODD dimensions measured with the eight symptoms referred to in the DSM-IV-TR (American Psychiatric Association, 2000) across boys and girls and across parents' and teachers' reports.

Measurement invariance analyses follow several sequential steps (e.g., Vandenberg & Lance, 2000). As a starting point, configural invariance is supported when the same symptoms are used to classify a construct, implying that the pattern of zero and nonzero loadings is similar across groups. First, invariance of factor loadings (metric invariance or weak measurement invariance) implies that the constructs themselves are the same, and is particularly important both in terms of relating factors to other constructs for different groups with cross-sectional data and for evaluating patterns of relations among variables in the same group over time with longitudinal data (Marsh, Nagengast, & Morin, 2013). This means that the corresponding factors have the same meaning in the different groups, i.e., the strength of the relations between each symptom and its ODD dimension is the same for both sexes and/or for parents and teachers. Invariance of factor loading is sufficient for evaluating relations among variables or relating ODD factors to other constructs. Second, invariance of item thresholds/intercepts (scalar invariance) implies that differences between items' mean levels in the groups of responses considered can be explained in terms of differences at the latent factor mean levels. Hence, strong measurement invariance (metric plus scalar) provides a justification for the interpretation of response-group differences based on latent means (Marsh, et al., 2013). Only if scalar invariance is achieved can ODD scores be meaningfully compared across sexes/informants. Third, equivalence of item residual variances or uniqueness (strict measurement invariance) tests whether the amount of item variance not accounted for by the factor is the same across groups in each item. It is a prerequisite for comparing observed or factor ODD scores that do not control for measurement error. Jointly with equivalence of factor variances, it is a proper test of invariant reliabilities for ODD dimensions. Finally, structural invariance tests the equivalence of structural parameters: factor variances (dispersion of the latent variables or variability of the construct, i.e., equivalent ranges of ODD continuum dimensions), factor covariances (relation between factors, i.e., constant conceptual ODD domain), and latent means (such as more traditional analyses with ANOVA or t-test).

Given that, in child psychopathology, and especially in the field of childhood disruptive behaviour problems, it is recommended to obtain information from several reporters (Hunsley & Mash, 2007), and that lack of agreement between informants tends to be the rule (de los Reyes & Kazdin, 2005), it is necessary to study whether there is equivalence in the measurement of the dimensions across informants. As mentioned earlier, measurement invariance is a prerequisite for subsequent valid mean comparisons, a task routinely performed in research work. Furthermore, clinical and research work commonly includes populations of both boys and girls. Therefore, we aimed to examine the invariance of the dimensions across sex. The goal of this work, then, is to evaluate the measurement equivalence of Burke’s model for ODD symptom dimensions across sex and informant (parents and teachers) in a community sample of preschool children.

Method

Participants

The data are from the first assessment of a large-scale longitudinal study of behavioural problems in preschool children from age 3. Details of the sampling procedures are described in (Ezpeleta, et al., 2012). Briefly, a cross-sectional two-phase design began with the selection of a random sample of 2,283 children from the census of preschoolers in grade P3 (3-year-olds) in Barcelona (Catalonia, Spain). A total of 1,341 families (58.7%) agreed to participate in the first phase. The parents of children participating in this first phase completed the *Strengths and Difficulties Questionnaire* parents' version (see below), which was used for screening purposes.

In the second phase, all children with a positive screening for behavioural problems and a random sample of 30% of children with a negative screening were invited to continue. The final second phase sample included 622 families (10.6% of those invited refused to participate in the second phase) and 94 teachers from 54 schools. No differences were found on comparing participants and refusals by sex (*p* = .82) or by type of school (*p* = .85). Children's mean age was 3.0 (*SD* = 0.16), 311 were boys (50.0%) and 89.5% were white, while 33.8% were of high socioeconomic status, 44.9% middle, and 21.3% low. Weighted DSM-IV prevalences in the sample, based on the *Diagnostic Interview for Children and Adolescents for Parents of Preschool and Young Children* (DICA-PPYC; Ezpeleta, Osa, Granero, Doménech, & Reich, 2011), were as follows: 3.7% of the children presented attention deficit/hyperactivity disorder, 6.9% ODD, 1.4% conduct disorder, 0.4% major depression, 2.2% separation anxiety, 3.7% specific phobia, and 1.9% social phobia.

Instruments

The ODD symptoms scores were obtained through four items of the *Strengths and Difficulties Questionnaire* (SDQ$^{3-4}$; Goodman, 1997) scale for conduct problems related to DSM-IV-TR (American Psychiatric Association, 2000) ODD symptoms ("Often has temper tantrums or hot tempers", "Often argumentative with adults", "Generally obedient, usually does what adults request", "Can be spiteful to others"), plus four items from the DSM-IV-TR definition of ODD not included in the questionnaire but added to the list of questions with the same response format (“Often deliberately annoys others”, “Often blames others for his/her mistakes or bad behaviour”, “Is easily offended by things others say”, “Is often angry and resentful”) were used for the analyses of invariance of the ODD dimensions. The items have three response options (0: *not true*; 1: *somewhat true*; 2: *certainly true*). Reverse items were coded in the direction of higher scores indicating more psychopathology. Parents (*N* = 622) and teachers (*N* = 615) answered the ODD questions. Each of the 88 teachers participating rated between 1 and 20 children (*Mdn* = 7).

Procedure

The longitudinal project was approved by the ethics review committee of the authors’ institution. Heads of the participating schools and parents were provided with a full description of the study. Families were recruited at the schools and gave written consent. All parents of children from grade P3 at the participating schools were invited to answer the SDQ[3-4], which was completed by families at home and returned to the schools, and were interviewed at the school. After obtaining consent from the parents, the questionnaire was given to the teachers for completion before the end of the academic year.

Statistical Analysis

The statistical analyses were carried out with Mplus7 (Muthén & Muthén, 1998-2012). Given the multistage sample, data corresponding to the second phase were analyzed with the case weighting procedure, with sampling weights inversely proportional to the probability of participant selection. Confirmatory Factor Analysis (CFA) was conducted using Weighted Least Squares Means and Variance (WLSMV) adjusted for the categorical data method of estimation. As long as the items are categorical (three response options), the distribution of each response scale of the items is replaced by a continuous distribution, having a probability curve derived from the normal distribution. Therefore, the three response categories representing a percentage of the sample are replaced by two thresholds in the normal distribution. Goodness-of-fit was assessed with the common fit indices (Jackson, Gillaspy, & Purc-Stephenson, 2009): χ^2, comparative fit index (CFI), and Root Mean Square Error of Approximation (RMSEA).

Burke’s model consists of an 8-item and 3-factor model, Symptoms 6-7-8 loading on negative affect, Symptoms 1-2-3 on oppositional behaviour, and Symptoms 4-5 on antagonistic behaviour. First, the model fit for baseline models in each sex separately and initial configural models across sex (multi-group approach) within teachers’ and parents’ responses was examined.

Second, invariance across sex was measured. Table 1 shows model identification for each step of the invariance analysis (Byrne, 2012), comparing progressively more constrained nested models (from least to most restrictive), and following the common sequence (Millsap & Yun-Tein, 2004; Vandenberg & Lance, 2000). We used the factor-variance strategy or fixed-factor method rather than the marker-variable strategy or reference-variable method, because the non-invariance of the reference variable when an anchor item is used is likely to cause severe Type I error inflation by forcing the unequal parameters to be invariant across groups (Byrne, 2012; Kim & Yoon, 2011). Theta parameterization was used, so that residual variances are allowed to be parameters in the model and strict measurement invariance can be tested (Kim & Yoon, 2011; Muthén & Muthén, 1998-2012). Design-based multilevel CFA strategy (i.e., the Type = COMPLEX routine in MPlus) was used for teachers' responses, to account for the hierarchical data structure due to cluster sampling, by specifying one single model for each group and then adjusting the overall model chi-square value and the standard errors of the parameter estimates with respect to the degree of data dependency (Kim, Kwok, & Yoon, 2012).

And third, measurement invariance across informants was assessed, considering the responses of teachers and parents as repeated measures, with a single-sample approach to account for non-independence of the observations; thus, error covariances between analogous items were also freely estimated (Ferrando, 2000), in addition to factor covariances. Regarding measurement invariance analyses, the same sequence and series of constraints as in the multi-group approach were considered across teachers' and parents' ratings (Table 1).

For both analyses, when full invariance was not achieved, we examined the fit indices of partially invariant models in which parameters of one item were relaxed sequentially with a backward procedure (Kim & Yoon, 2011). The α level for testing nested models with the scaled chi-square difference (Bryant & Satorra, 2012) was set at .01 (e.g., Dekovic, et al., 2006; Ferrando, 2000; Gomez, 2013) for Type I error control (Green & Babyak, 1997). Internal consistency of the dimensions was measured through the omega coefficient (McDonald, 1999).

Results

Table 2 shows the results of CFAs across sex within teachers' (top) and parents' (centre) responses. Baseline models for each sex (T0a, T0b, P0a and P0b in Table 2) and configural invariance across sex for both informants (T1 and P1 in Table 2) was supported, since model fit was satisfactory (CFI ≥ .96; RMSEA ≤ .069). Thus, the 3-factor model proved to be a good solution for both parents and teachers as informants in both sexes. Complete measurement and structural invariance was found, indicating that all parameters were equivalent across girls and boys within both types of informant. Moreover, given that full strong invariance (equivalence of factor loadings and item thresholds) was achieved, comparison of latent means could be conducted (Vandenberg & Lance, 2000), and the latent means for all the factors were found to be equivalent across sex (T6 and P6 in Table 2).

Model fit for these final constrained models across sex was also satisfactory (CFI ≥ .98; RMSEA ≤ .042).

Because support was found for complete invariance across sex, internal consistency and repeated-measure CFAs across informants were conducted across all respondents, girls and boys jointly. Internal consistency (omega coefficient) for teachers' and parents' responses was, respectively, .85 and .68 for negative affect, .79 and .70 for oppositional behaviour, and .81 and .53 for antagonistic behaviour.

Table 2 (bottom) also shows model fit for baseline models in each informant separately (T0 and P0 in Table 2) and the results of the repeated-measure CFA across informants. Full metric invariance (equivalence of factor loading) was obtained (TP2 in Table 2), whereas full strong invariance (adding equivalence of item thresholds) was not (TP3 in Table 2). Partial strong invariance was not achieved either, since only 11 of the 16 item thresholds (less than 80%; Dimitrov, 2010) were invariant (equivalent) across informants (TP3a in Table 2, in bold). Model fit for this final constrained model across informants was satisfactory (CFI = .97; RMSEA = .036). Standardized parameters can be seen in Figure 1. The five threshold parameters showing non-invariance were the first for “loses temper” and “argues with adults”, and the second for “spiteful/vindictive”, “defies people” and “annoys people”: three were higher for teachers' than for parents' ratings, while two were in the opposite direction. Given that threshold parameters can be transformed into *z*-values, this shows that teachers rated “loses temper” and “argues with adults” more frequently as “*not true*” (higher first threshold) and “spiteful/vindictive” less frequently as “*certainly true*“ (higher second threshold, *z*-value corresponding to the accumulated percentage of “*not true*” and “*somewhat true*” options) than parents. By contrast, teachers rated “defies adults” and “annoys people” (lower second threshold) more frequently as “*certainly true*” than parents did.

Factor correlations between teachers' and parents' responses for analogous factor pairs were *r* = .09 (*p* = .250) for negative affect, *r* = .31 (*p* < .001) for oppositional behaviour, and *r* = .30 (*p* < .001) for antagonistic behaviour.

Discussion

ODD dimensions (negative affect, oppositional behaviour, and antagonistic behaviour) identified in preschool children performed in the same way in boys and girls, since all items showed strong measurement invariance (i.e., none of them showed differential item functioning) across sex. In addition, latent means did not differ between boys and girls for either parents' or teachers' reports considered separately. However, when examining measurement invariance between informants, some differences were found, which indicate that the equivalence of the ratings of parents and teachers is not complete, since given the same underlying level of the latent trait, one informant provides different item responses from those of the other.

Parents' and teachers' factor loadings were fully equivalent (metric invariance), but scalar invariance was not attained, because only the symptoms "touchy/ annoyed", "blames others" and "angry/resentful" were fully invariant across both types of informant. Given that the last category of the ODD items ("*certainly true*") is by far the least endorsed, the main interest for scalar invariance may focus on the first threshold parameter (percentage for "*not true*"), which was found to be higher for teachers' than for parents' responses in two of the oppositional behaviour items ("loses temper" and "argues with adults"). Thus, parents may tend to rate these symptoms higher than teachers, given the same latent trait level.

The lack of invariance in the dimensions has several clinical and research implications. The absence of equivalence in the item thresholds means that comparisons of observed means from parents and teachers are not readily interpretable. If, for instance, we wish to study which dimension (negative affect/oppositional/ antagonistic) most improves through a treatment for ODD with a pre-post design in preschoolers, we cannot treat jointly parents' and teachers' scores on the dimensions for those analyses that involve observed means, such as direct comparisons of change scores between informants. Furthermore, we cannot calculate absolute parent-teacher agreement because it is based on systematic differences in mean scores (whether ratings from both types of informant resemble one another or not); we can only calculate Pearson correlation coefficients (as factor correlations between analogous dimensions), which merely consider the ordering of the children as scored by each informant. The absence of scalar equivalence between parents and teachers reports may also have implications when a cut-off score is set for the classification of children's ODD behaviours. In this case, as it was observed that parents tend to rate ODD behaviours as more severe than teachers, it would be necessary to use different cut-off points for each informant. However, invariance of factor loadings is sufficient for evaluating the relations among variables or relating ODD factors to other constructs, such as those obtained through studies involving convergent validity, prediction, or comorbidity.

The relative agreement between parents and teachers (factor correlations) on dimension scores was better (but still low-level) for dimensions describing overt behaviours (oppositional/antagonistic behaviour) than that for dimensions describing mood (negative affect), which was very low (de los Reyes & Kazdin, 2005). This lack of agreement may be a reflection of the greater difficulties for identifying the mood state (negative affect) than the behavioural consequences (opposition). Internal consistency for parents was lower than that found for teachers.

If the constructs themselves are perceived in essentially the same way by parents and teachers (metric invariance), then the non-equivalences observed in item thresholds (scalar invariance) might be associated with cross-context differences in children's behaviour that is rated differently by parents and teachers. The discrepancies observed between parents and teachers might be attributable, in part, to differences in the context or the situation where the child behaves (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). The school setting is more structured than the home, and this could lead to more ODD symptoms at home than at school (Drabick, Gadow, & Loney, 2007). An alternative explanation is that teachers, who have experience with many children, have a better framework for evaluating what is normative at this age in comparison to parents. In general, the levels of internalizing and externalizing problems reported by teachers are lower than those reported by parents (Munkvold, Lundervold, Lie, & Manger, 2009; van der Ende & Verhulst, 2005). There is some debate about the usefulness of each informant, some studies reporting that parent reports are more predictive of psychological outcomes (Ferdinand, van der Ende, & Verhulst, 2007), and others the opposite. For the case of ODD, Drabick, Bubier, Chen, Price, and Lanza (2011) confirmed the importance of including teachers' reports for the assessment of this disorder, and found that teacher-reported symptoms were more predictive of psychopathological outcomes than parent-reported ones. Based on this area of research, recent literature indicates that child psychopathology, and specifically ODD, must be conceptualized as source-specific phenomena: different groups of children, with different characteristics, are identified depending on the informant and on how the information is combined (Drabick, et al., 2007; Munkvold, et al., 2009). And, in this line, the topic of disagreements between informants in the assessment process is receiving a great deal of attention, changing the view of disagreements as a source of unreliability to one whereby they are viewed as a source of meaningful information about the clinical picture of the child (De Los Reyes, 2011). Assessment, classification and treatment are affected by informant disagreements (De Los Reyes & Kazdin, 2005). Discrepancies between reporters reveal important information about the children's behavioural expression, which highlights the need to collate information from various contexts (De Los Reyes, 2011). Furthermore, discrepancies may indicate a different prognosis (Dirks, et al., 2012), and may have an effect in the context of treatment planning, permitting – when the differences are well understood – the building of therapeutic alliances, and facilitating the identification of treatment targets and the design of interventions that take into account the different perceptions of the problem (Achenbach, 2011). Therefore, and in line with this point of view, the absence of full scalar invariance might also be interpreted as informative.

To our knowledge, this is the first study of the invariance of ODD dimensions in preschool children. Previous work with the whole construct of ODD, as assessed in disruptive behaviour questionnaires, had shown equivalence of items loadings of mothers' and fathers' reports across children in different countries (Thailand, Brazil, North America, Australia, Malaysia) (Burns, et al., 2008; Burns, Desmul, Walsh, Silpakit, & Ussahawanitchakit, 2009) and across ages 9 to 16 (Sterba, et al., 2010). Measurement invariance across sex has also been demonstrated for American and Malaysian children (and boys scored higher than girls) (Burns, Walsh, Gomez, & Hafetz, 2006).

Some limitations should be taken into account when interpreting the present results. We recruited cases from a general population, resulting in a response rate of 59%; even so, given the purpose of the study, which was to provide evidence on measurement invariance of ODD dimensions, the participation rate might not adversely affect the results. It should also be mentioned that few families of low socioeconomic status participated, and this must be considered for generalization purposes.

To summarize, measurement invariance is a sound way of testing if we can compare the means across different groups, occasions, or situations, and hence, if there is a basis to draw scientific inferences from the measures obtained (Meredith, 1993). Measurement invariance results inform us about whether, under different conditions of measuring ODD dimensions (boys-girls, parents-teachers), measurements yield measures of the same attributes. We can conclude that ODD dimensions derived from boys and girls are fully equivalent and comparable: there is no differential item functioning when ODD dimensions are assessed across sex. Also, the discrimination (metric invariance) of the ODD items studied functions in the same way for parents and teachers. However, accepting the equality of the construct, the informants (parents and teachers) score differently, and consequently, the practical implication is that mean scores provided by these reporters might not be compared. According to our results, the latent ODD dimensions are similarly conceptualized by parents and teachers (i.e., parents' and teachers' ODD symptoms are analogously associated with the constructs of negative affect, oppositional behaviour, and antagonistic behaviour), but they differ in the level of the behaviours observed (they score them differently, parents rating some symptoms higher). The lack of full or partial scalar invariance supports the concept of ODD as a source-specific disorder, which is also the view of other authors (Drabick, et al., 2007; Gadow & Drabick, 2012; Strickland, Hopkins, & Keenan, 2012), and highlights the need of including both reporters (parent and teachers) in the assessment process of children with ODD. Further research may confirm whether scalar equivalence is maintained or not in older individuals.

References

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology, 40*, 80-86.

American Psychiatric Association. (2000). *DSM-IV Diagnostic and statistical manual of mental disorders* (4th Text Revised ed.). Washington, DC: American Psychiatric Press.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference χ^2 testing. *Structural Equation Modeling, 19*, 372-398.

Bufferd, S. J., Dougherty, L. R., Carlson, G. A., & Klein, D. N. (2011). Parent-reported mental health in preschoolers: findings using a diagnostic interview. *Comprehensive Psychiatry, 52*, 359-369.

Burke, J., & Loeber, R. (2010). Oppositional defiant disorder and the explanation of the comorbidity between behavioral disorders and depression. *Clinical Psychology: Science and Practice, 17*, 319-326.

Burke, J. D., Hipwell, A. E., & Loeber, R. (2010). Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child Adolescent Psychiatry, 49*, 484-492.

Burns, G. L., de 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.

Burns, G. L., Walsh, J. A., Gomez, R., & Hafetz, N. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment, 18*, 452-457.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Taylor and Francis Group.

De Los Reyes, A. (2011). More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 40,* 1-9

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.

Dekovic, M., ten Have, M., Vollebergh, W., Pels, T., Oosterwegel, A., Wissink, I., ... Ormel, J. (2006). The cross-cultural equivalence of parental rearing measure: EMBU-C. *European Journal of Psychological Assessment, 22*, 85-91.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121-149.

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Embracing not erasing contextual variability in children's behavior - theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry, 53*, 558-574.

Drabick, D. A. G., Bubier, J., Chen, D., Price, J., & Lanza, H. I. (2011). Source-specific 0ppositional defiant disorder among inner-city children: Prospective prediction and moderation. *Journal of Clinical Child and Adolescent Psychology, 40,* 23-35.

Drabick, D. A. G., Gadow, K. D., & Loney, J. (2007). Source-specific oppositional defiant disorder: Comorbidity and risk factors in referred elementary schoolboys. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 92-101.

Ezpeleta, L., Granero, R., Osa, N. d. l., Penelo, E., & Doménech, J. M. (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. *Journal of Child Psychology and Psychiatry, 53*, 1128-1138.

Ezpeleta, L., Osa, N. d. l., Granero, R., Doménech, J. M., & Reich, W. (2011). The Diagnostic Interview for Children and Adolescents for Parents of Preschool Children. *Psychiatry Research, 190*, 137-144.

Ferdinand, R. F., van der Ende, J., & Verhulst, F. C. (2007). Parent-teacher disagreement regarding behavioral and emotional problems in referred children is not a risk factor for poor outcome. *European Child & Adolescent Psychiatry, 16*, 121-127.

Ferrando, P. J. (2000). Testing the equivalence among different item response formats in personality measurement: A structural equation modeling approach. *Structural Equation Modeling, 7*, 271-286.

Gadow, K. D., & Drabick, D. A. G. (2012). Anger and irritability symptoms among youth with ODD: Cross-informant versus source-exclusive syndromes. *Journal of Abnormal Child Psychology, 40,* 1073-1085.

Gomez, R. (2013). DSM-IV ADHD symptoms self-ratings by adolescents: Test of invariance across gender. *Journal of Attention Disorders, 17*, 3-10.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581-586.

Green, S. B., & Babyak, M. A. (1997). Control of Type I errors with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research, 32*, 39-51.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment *Annual Review of Clinical*

Psychology, 3, 29-51.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14,* 6-23.

Kim, S. H., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*, 212-228.

Kim, E. S., Kwok, O.-m., & Yoon, M. (2012). Testing factorial invariance in multilevel data: A Monte Carlo study. *Structural Equation Modeling, 19*, 250-267.

Lavigne, J. V., Cicchetti, C., Gibbons, R. D., Binns, H. J., Larsen, L., & DeVito, C. (2001). Oppositional defiant disorder with onset in preschool years: Longitudinal stability and pathways to other disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1393-1400.

Lavigne, J. V., Lebailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child and Adolescent Psychology, 38*, 315-328.

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and La Dolce Vita effects. *Developmental Psychology, 49,* 1194-1218.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale: Erlbaum.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525-543.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39,* 479-511.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Munkvold, L., Lundervold, A., Lie, S. A., & Manger, T. (2009). Should there be separate parent and teacher-based categories of ODD? Evidence from a general population. *Journal of Child Psychology and Psychiatry*, *50*, 1264-1272.

Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology, 119*, 726-738.

Sterba, S. K., Copeland, W., Egger, H. L., Costello, E. J., Erkanli, A., & Angold, A. (2010). Longitudinal dimensionality of adolescent psychopathology: Testing the differentiation hypothesis. *Journal of Child Psychology and Psychiatry, 51,* 871-884.

Strickland, J., Hopkins, J., & Keenan, K. (2012). Mother-teacher agreement on preschoolers' symptoms of ODD and CD: Does context matter? *Journal of Abnormal Child Psychology, 40*, 933-943.

Stringaris, A., & Goodman, R. (2009). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry, 50*, 216-223.

van der Ende, J., & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child & Adolescent Psychiatry, 14*, 117-126.

Vandenberg, R., & Lance, C. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Table 1. *Model identification and sequential steps taken for invariance analysis.*

Step	Model identification	Parameters for both 1^{st} and 2^{nd} groups
1	Configural model: Equal form *	Factor loadings (λ) free to vary Item thresholds (τ) free to vary Uniquenesses (δ) fixed at 1 ** Factor variances (φ_i) fixed at 1 Factor covariances (φ_{ij}) free to vary Latent means (κ) fixed at 0
	Tests of invariance	Sequential constraints in 2^{nd} group
2	Factor loadings: Weak (metric)	**λ fixed to be equal to 1^{st} group** τ free to vary δ fixed at 1 (as in 1^{st} group) ** **φ_i free to vary** φ_{ij} free to vary κ fixed at 0 (as in 1^{st} group)
3	Item thresholds: Strong (metric + scalar)	λ fixed to be equal to 1^{st} group **τ fixed to be equal to 1^{st} group** δ fixed at 1 (as in 1^{st} group) ** φ_i free to vary φ_{ij} free to vary **κ free to vary**
4	Item residual variances or uniquenesses: Strict (strong + uniquenesses) ***	λ fixed to be equal to 1^{st} group τ fixed to be equal to 1^{st} group **δ free to vary** φ_i free to vary φ_{ij} free to vary κ free to vary
5	Factor variances and covariances	λ fixed to be equal to 1^{st} group τ fixed to be equal to 1^{st} group δ fixed at 1 to be equal to 1^{st} group **φ_i fixed at 1 to be equal to 1^{st} group** **φ_{ij} fixed to be equal to 1^{st} group** κ free to vary
6	Latent means	λ fixed to be equal to 1^{st} group τ fixed to be equal to 1^{st} group δ fixed at 1 to be equal to 1^{st} group φ_i fixed at 1 to be equal to 1^{st} group φ_{ij} fixed to be equal to 1^{st} group **κ fixed at 0 to be equal to 1^{st} group**

Note: Steps 2-4 for measurement invariance; steps 5-6 for structural invariance. In bold: specific changes at each step, with respect to the immediately previous step.

* To avoid the use of a marker item (Kim & Yoon, 2011) (MPlus default for factor loadings), the factor loadings and item thresholds of the first item for each factor were also freely estimated, but all factor variances were fixed at 1 and all latent means were fixed at 0 (Byrne, 2012; Muthén & Muthén, 1998-2012).

** When a factor loading and an item threshold for a categorical factor indicator are free across groups, the residual variance/uniqueness for the variable must be fixed at 1 for identification purposes (Muthén & Muthén, 1998-2012).

*** Test for equivalence of residual variances/uniquenesses proceeds backwards: item residual variances (which were fixed at 1 in all groups in the previous step 3) are freely estimated in the second group and then compared to the previous model in which all uniquenesses had been fixed at 1 (see, for example, http://psych.unl.edu/psycrs/948_2011/13a_Invariance_in_IRT-IFA.pdf).

Table 2. *Fit indices for measurement and structural invariance analysis across sex within teachers' (T) and parents' (P) responses (top and centre) and repeated-measures measurement invariance analysis across teachers' and parents' responses (bottom).*

Informant	Model	Goodness-of-fit indices			Comparison of nested models		
		χ^2 (df)	CFI	RMSEA	Models compared	$\Delta\chi^2$ (Δdf)	*p*
Model fit and invariance across sex within each informant							
Teachers	T0a: females	37.137 (17)	.987	.062			
	T0b: males	42.323 (17)	.976	.069			
	T1: same configuration (equal form)	79.117 (34)	.983	.066			
	T2: weak invariance (equal loadings)	73.556 (39)	.987	.054	T2 vs. T1	5.171 (5)	.395
	T3: strong invariance (plus equal thresholds)	92.725 (52)	.984	.050	T3 vs. T2	22.265 (13)	.051
	T4: strict invariance (uniquenesses free)*	86.369 (44)	.984	.056	T3 vs. T4*	11.374 (8)	.181
	T5: plus equal factor variances-covariances	86.976 (58)	.989	.040	T5 vs. T3	6.354 (6)	.385
	T6: plus equal means	94.840 (61)	.987	.042	T6 vs. T5	7.663 (3)	.054
Parents	P0a: females	25.287 (17)	.982	.040			
	P0b males	36.033 (17)	.965	.060			
	P1: same configuration (equal form)	61.122 (34)	.973	.051			
	P2: weak invariance (equal loadings)	61.366 (39)	.977	.043	P2 vs. P1	4.694 (5)	.454
	P3: strong invariance (plus equal thresholds)	79.251 (52)	.973	.041	P3 vs. P2	18.646 (13)	.135
	P4: strict invariance (uniquenesses free)*	68.182 (44)	.976	.042	P3 vs. P4*	11.979 (8)	.152
	P5: plus equal factor variances-covariances	72.844 (58)	.985	.029	P5 vs. P3	3.650 (6)	.724
	P6: plus equal means	71.833 (61)	.989	.024	P6 vs. P5	1.550 (3)	.671
Model fit and invariance across informants							
T0: teachers		67.416 (17)	.980	.069			
P0: parents		43.245 (17)	.973	.050			
TP1: same configuration (equal form)		152.322 (81)	.973	.038			
TP2: weak invariance (equal loadings)		154.132 (86)	.975	.036	TP2 vs. TP1	7.929 (5)	.160
TP3: strong invariance (TP2 plus equal thresholds)		238.148 (99)	.948	.048	TP3 vs. TP2	127.72 (13)	<.001
TP3a: TP2 (all λ equal) plus 11 τ equal		**167.744 (94)**	**.972**	**.036**	TP3a vs. TP2	18.509 (8)	.018

Note. * Test for invariance of residual variances/uniqueness proceeds backwards: uniquenesses are first freely estimated in the second group (Model #4), and are then compared to the model in which all uniquenesses are fixed at 1 in the second group so as to be in line with the first group (Model #3).

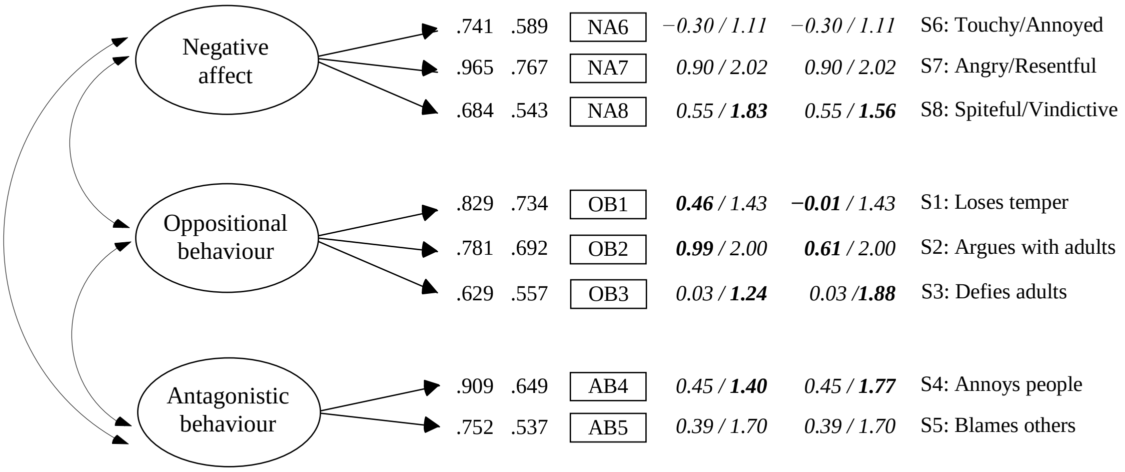
Negative affect
Oppositional behaviour
Antagonistic behaviour
.741 .589 NA6 −0.30 / 1.11 −0.30 / 1.11 S6: Touchy/Annoyed
.965 .767 NA7 0.90 / 2.02 0.90 / 2.02 S7: Angry/Resentful
.684 .543 NA8 0.55 / 1.83 0.55 / 1.56 S8: Spiteful/Vindictive
.829 .734 OB1 0.46 / 1.43 −0.01 / 1.43 S1: Loses temper
.781 .692 OB2 0.99 / 2.00 0.61 / 2.00 S2: Argues with adults
.629 .557 OB3 0.03 / 1.24 0.03 /1.88 S3: Defies adults
.909 .649 AB4 0.45 / 1.40 0.45 / 1.77 S4: Annoys people
.752 .537 AB5 0.39 / 1.70 0.39 / 1.70 S5: Blames others

Figure 1. Burke’s final model TP3a (8-item and 3-factor): Standardized factor loadings (λ; normal font) and item thresholds (τ_1/τ_2; italics) across teachers’ (left) and parents’ (right) responses. In bold: Parameters non-equivalent across informants.