COPYRIGHT NOTICE

Abstract

This study used confirmatory factor analysis (CFA) to determine the best model for Oppositional Defiant Disorder (ODD) symptoms in children aged 3 to 15 years, as presented in the Disruptive Behavior Rating Scale. Teachers' ratings of the ODD symptoms of 213 children from the general school community were obtained. The findings provided most support for a bifactor model based on Stringaris and Goodman's (2009) three-factor model (primary factors for irritable, hurtful, and headstrong). The general factor, but not the specific factors in the model, showed high omega hierarchical and explained common variance. Thus, only the general factor in this model can be meaningfully interpreted. Also, the general factor was supported with regard to external validity. Specifically, this factor, but not the specific factors, correlated strongly with ADHD inattention and hyperactivity/impulsivity symptom groups, and other measures of behavioural and emotional problems. The taxonomic, diagnostic, practical, and research implications of the findings are discussed.

Keywords: Oppositional Defiant Disorder; confirmatory factor analysis; community sample; teacher ratings, bifactor model.

Factor structure of Teacher Ratings of the ODD Symptoms in Children

Oppositional Defiant Disorder (ODD) is a common childhood disorder (American Psychiatric Association [APA], 2000; APA, 2013), and refers to a recurrent pattern of negativistic, defiant, disobedient, and hostile behavior toward authority figures that persists for at least six months (APA, 2013). For diagnosis, the DSM-IV (APA, 1994) and the DSM-IV-TR (APA, 2000) have eight ODD symptoms (see Figure 1 for description of symptoms) organised together under one group, thereby implying support for a unidimensional or one-factor model for these symptoms. With some minor changes to the wording, the DSM-5 has the same eight symptoms, but they are placed into three symptom groups: anger/irritable (comprising symptoms for temper, anger, and touchy), vindictiveness (comprising the symptom for spiteful), and argumentative/defiant behaviour (comprising symptoms of arguing, annoying, defiance, and blaming). Although the three symptom groups in the DSM-5 hint at the possibility that ODD might be multidimensional, it is still viewed in categorical terms, with the disorder being either present or absent based on functional impairment and the presence of at least four of the eight symptoms.

The three ODD symptom groups in the DSM-5 correspond to an earlier, a priori, three-dimensional model of ODD proposed by Stringaris and Goodman (2009). This model is comprised of the dimensions of irritable, hurtful, and headstrong (see Figure 1 for the symptoms within these dimensions). These dimensions are the same as the anger/irritable, vindictiveness, and argumentative/defiant behaviour dimensions, respectively, in the DSM-5. Confirmatory Factor Analysis (CFA) studies have supported the Stringaris and Goodman (2009) model (e.g., Krieger et al., 2013).

To date, several other factor models have been proposed for ODD. Aebi et al. (2010) have found support for a different three-factor model. The difference between their model and Stringaris and Goodman's model is that the symptom for 'annoy' falls within the hurtful factor, and not in the headstrong factor (see Figure 1). Burke, Waldman, and Lahey (2010) proposed yet another three-factor model, with factors for negative affect, oppositional behaviour, and antagonistic behaviour (see Figure 1 for the symptoms within these dimensions). As shown in Figure 1, the negative affect construct has the symptoms of touchy and angry that are part of the irritable construct proposed by Stringaris and Goodman (2009) and Rowe, Costello, Angold, Copeland, and Maughan (2010). Thus, negative affect and irritable are highly comparable constructs (Burke, 2012; Burke et al., 2010).

Besides three-factor models, two-factor models have also been supported by CFA studies. Rowe et al. (2010) reported factors for irritable and headstrong/spiteful (see Figure 1 for the symptoms within these dimensions). The symptoms within the irritable factor were identical to the symptoms for the irritable factor in Stringaris and Goodman's (2009) three-factor model. In another model, Burke, Loeber, Lahey, and Rathouz (2005) reported factors for oppositional behavior and negative affect (see Figure 1 for the symptoms within these dimensions). However, this model is limited as it does not include two of the eight ODD symptoms (blames others and annoys others). A more recent study by Burke et al. (2014) found support for a bifactor model of ODD. In this model, all the symptoms loaded on a general ODD factor, and specific factors after accounting for their variances in the general factor. The specific factors were irritable (comprising touchy, angry, and temper symptoms) and oppositional (the remaining five symptoms). It should be noted however that, conceptually, the specific factors in a bifactor model are not the same as the first-order factors found in the corresponding first-order factor model. The

specific factors represent common variance after controlling for the general factor, whereas first-order factors include variances that are part of the general ODD factor.

A few studies have compared the relative support for the different ODD factor models. For example, Krieger et al. (2013) found better fit for Stringaris and Goodman's (2009) three-factor model over the one-factor model, and Rowe et al.'s (2010) two-factor model. Lavigne, Bryant, Hopkins, Karen, and Gouze (2015) found better fit for Burke et al.'s (2005) two-factor model over the one-factor model, Burke et al.'s (2010) three-factor model, and Rowe et al.'s (2010) two-factor model. Herzhoff and Tackett (2016) reported better fit for Burke et al.'s (2005) two-factor model over the three-factor models proposed by Aebi et al. (2010), Burke et al. (2010), and Stringaris and Goodman (2009), and the two-factor model proposed by Rowe et al. (2010). Ezpeleta, Granero, de la Osa, Penelo, and Domenech (2012) found more support for Burke et al.'s (2010) three-factor model than for Burke et al.'s (2005) two-factor model, Stringaris and Goodman's (2009) three-factor model, and the one-factor model. More recently, Burke et al. (2014) found that their bifactor model showed better fit than other two- and three-factor models.

Studies have also examined support for the external validity (convergent and divergent) of the factors in two- and three-factor CFA models. In relation to the three-factor CFA models, Stringaris and Goodman (2009) found that the irritable factor was associated with emotional disorders. Although a cross-sectional analysis showed associations between all three factors and conduct problems, the hurtful factor showed differential associations with aggression, conduct problems, and callous–unemotional behaviors three years later. Aebi et al. (2010) reported that the irritable, hurtful, and headstrong factors in their ODD model were positively associated with conduct problems, and negatively associated with the prosocial behavior factors of the Strengths

and Difficulties Questionnaire (SDQ; Goodman, 1997). Additionally, the irritable factor showed a positive association with the emotional symptoms factor of the SDQ. Like Stringaris and Goodman's (2009) and Rowe et al.'s (2010) irritable construct, the negative affect constructs in Burke et al's (2010) model showed associations with depression and anxiety. Negative affect was also associated with neuroticism and parental reports of mood and personality difficulties as young adults. Krieger et al. (2013) found that, for their two-factor model, the irritable factor had a strong association with childhood emotional disorders and a history of maternal depression and suicidality, while the headstrong factor was uniquely associated with childhood ADHD and maternal history of ADHD. The hurtful factor was associated with childhood conduct disorder. For Rowe et al.'s (2010) two-factor model, Herzhoff and Tackett (2016) demonstrated convergent and divergent validity for the irritability and headstrong/spiteful factors in terms of differential associations with later behavioural problems and personality.

In summary, while three-factor, two-factor, and bifactor factor models have been proposed and supported for ODD, and there is some agreement for distinct factors for irritable/negative affect and headstrong/spiteful (or oppositional) factors, there are limitations with respect to existing data in this area. First, there is a lack of agreement on the best factor structures and the composition of the symptoms in the primary factors in these models. In terms of factor structure, there is some evidence that a bifactor conceptualization would be more preferable over one and oblique multidimensional first-order factor models, as shown in the study by Burke et al. (2014). The possibility of a bifactor model is also suggested by findings presented in past studies showing significant (and often high) correlations between the factors of multidimensional ODD models (Ezpeleta et al., 2012; Herzhoff & Tackett, 2016; Krieger et al., 2013). For example, the study by Ezpeleta et al. (2012) reported that the correlations of the

factors in Stringaris and Goodman's (2009) three-factor model were .48 between irritable and headstrong, .30 between irritable and hurtful, and .27 between headstrong and hurtful. For the Rowe et al. (2010) two-factor model, the correlation between the irritable and headstrong/spiteful factor was .51; and for Burke et al's (2010) three-factor model, the correlations were .37 between negative affect and oppositional, .34 between negative affect and antagonistic, and .37 between oppositional and antagonistic. Although Burke et al. (2014) examined a bifactor model, they did not examine bifactor models with specific factors aligned to previously proposed first-order multidimensional models. It is conceivable that such a model would turn out to be the optimum model.

A second limitation is that much of the data that has been used to examine the factor structure of ODD symptoms has been based on parent ratings and/or reports. Exceptions are the studies by Gomez (2016) and Ezpeleta et al. (2012) that have examined this for teacher ratings of the ODD symptoms. For teacher reports of preschool children, Ezpeleta et al. (2012) found more support for Stringaris and Goodman's (2009) three-factor model than Burke et al.'s (2010) three-factor model, Burke et al.'s (2005) two-factor model, and the one-factor model. Gomez (2016) examined the applicability of the one-factor model, Rowe et al.'s (2010) two-factor model, and three-factor models based on Aebi et al. (2010), Burke et al. (2010), and Stringaris and Goodman (2009) for teacher ratings of a group of Malaysian primary school children. Although the findings showed some support for all models examined, there was most support for Burke et al.'s (2010) three-factor model. Thus, based on the limited data available, it is not possible to infer the optimum ODD model for teacher ratings or reports.

As the DSM-5 views severity of ODD in terms of the presence of ODD across settings, and as teachers are useful sources of information for clinical diagnosis, it will be useful to know

the factor structure of ODD symptoms for children based on teacher reports. In this respect, it will be valuable to examine the applicability of previously proposed models, which are based on parent ratings of the ODD symptoms. In addition, it is conceivable that a bifactor conceptualization with specific factors aligned to one or more of the previously proposed first-order, multidimensional models would be supported. Such a model is possible as there are data showing high intercorrelations between ODD factors in past models, based on teacher ratings. Ezpeleta et al. (2012) reported that the correlations of the factors in Stringaris and Goodman's (2009) three-factor model were .58 between irritable and headstrong, .49 between irritable and hurtful, and .37 between headstrong and hurtful. For the Rowe et al. (2010) two-factor model, the correlation between the irritable and headstrong/spiteful factor was .64; and for Burke et al.'s (2010) three-factor model, the correlations were .52 between negative affect and oppositional, .53 between negative affect and antagonistic, and .59 between oppositional and antagonistic.

A third limitation is that, although there is support for the structure of Burke et al.'s (2014) model, the reliability and external validity of the factors in this model has not been examined. For a bifactor model it is possible to compute the explained common variance (ECV) and the omega hierarchical (ω_h; McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005) of the general and specific factors. The ECV of a general factor is the common variance explained by the general factor divided by the total common variance, and the ECV of a specific factor is the common variance explained by the specific factor divided by the total common variance. The ECV of the general factor will be high whenever there is little common variance beyond that of the general factor. Thus, high values indicate the presence of a general factor dimension in the bifactor model (Reise, Bonifay, & Haviland, 2013). The ω_h can be interpreted as an estimator of how much variance in summed (standardized) scores can be attributed to the general factor

(Brunner, Nagy, & Wilhelm, 2012). It is model-based, and is obtained by dividing the amount of trait variance explained by the general factor by the total amount of variance (trait plus error) explained by the general factor (and not the entire scale as in the case of ECV). The ω_h value for a specific factor [also referred to as omega-subscale (ω_s)] in a bifactor model can be computed by dividing the amount of specific variance (removing the variance that is part of the general factor) explained by the factor by the total amount of variance (trait plus general plus error) explained by that factor. The values for ω_h range from 0 to 1, with 0 indicating no reliability and 1 reflecting perfect reliability. According to Reise et al. (2013), ω_h values of at least .75 are preferred for meaningful interpretation of a scale. Overall, therefore, high ECV and ω_h (> .75) values indicate the presence of a general dimension in the bifactor model. For a first-order factor model, the comparable model-based reliability for is called omega (ω; McDonald, 1999). For this model, the ω value for a primary factor is computed by dividing the amount of trait variance explained by the factor by the total amount of variance (trait plus general plus error) explained by that factor. Thus, for the ODD symptoms in bifactor models, high ECV and ω_h (> .75) values would indicate the presence of a dominant ODD factor that would, in turn, justify the use of the total score from these measures.

Given existing limitations and omissions, the present study used CFA to examine the factor structure of DSM-IV/DSM-5 ODD symptoms, based on teacher ratings of children in the general school community. First, we examined one-factor (inspired by DSM-IV), two-factor (Rowe et al., 2010), and three-factor (Aebi et al., 2010; Burke et al., 2010; Stringaris & Goodman, 2009) models. We also examined bifactor models based on these two- and three-factor models, and the bifactor model proposed by Burke et al. (2014) (see Figure 1). Burke et al.'s (2005) two-factor model was not considered as it included only six of the eight ODD

symptoms. Secondly, we examined the external validities of the factors of the best of these models by examining the correlations of the factors in this model with teacher ratings of behavioral and emotional problems as presented in the SDQ (Goodman, 1997), and with teacher ratings of ADHD inattention (IA) and ADHD hyperactivity/impulsivity (HI) symptom groups. The SDQ has scales for Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems, and Prosocial Behavior. Based on our review of the literature, and the findings reported by Burke et al. (2014), we predicted that one of the bifactor models would show the best fit. Additionally, the general factor in this model would show relatively higher ECV and ω_h values, compared to the specific factors. In terms of the external validity of a bifactor model, we expected that, compared to the specific factors, the general factor would have stronger associations with all the SDQ behavioral and emotional problems, and IA and HI scores. Additionally, the specific factor for irritable/negative affect (if present in the optimum model) would have associations with SDQ emotional symptoms, and headstrong/oppositional (if present) would have some association with SDQ conduct problems.

Method

Participants

The sample was comprised of the teachers of 213 children (boys = 104, girls = 109). The mean age (*SD*) for all children was 8.52 (1.77) years. The children were from 10 Australian primary schools. Of the children, 91.2% had European background, 6.0% Asian, 2.4% Middle East/African, and 0.4% others (including indigenous Australian). These figures compare to around 90% European, 6.6% Asian, 1.2% of Middle East and African, and 2.2% others (including indigenous Australian) in the general Australian population (Australian Bureau of

Statistics, 2007). There was a close match in ethnicity between the Australian general population and the groups rated in the study, χ^2 (*df* = 3) = 1.68, *p* = *ns*; Yates χ^2 (*df* = 3) = 0.27, *p* = *ns*.

In terms of parental background relating to socioeconomic status, father's occupational status (or mother's when father's was not available) was coded according to the Australian Standard Classification of Occupations (ASCO; Australian Bureau of Statistics, 1997). This has nine major, hierarchically-organized occupational categories defined in terms of skills and specialization. In decreasing order, they are managers and administrators (coded 1), professionals (coded 2), associate professionals (coded 3), tradespersons (coded 4), advanced clerical and service workers (coded 5), intermediate clerical, sales and service workers (coded 6), intermediate production and transport workers (coded 7), elementary clerical, sales and service workers (coded 8), and labourers (coded 9). Those not employed were also coded 9 in this study. The mean occupational level for the parents of the children in this study was 4.61 (*SD* = 3.69). Overall, the occupational level was "middle-class".

Measures

Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998). The DBRS includes the DSM-IV symptoms (same as in DSM-IV-TR and DSM-5) for ADHD and ODD. Teacher ratings for the DBRS ODD symptoms were used for the study. For all symptoms, the word "often" was excluded. Respondents rate each symptom on a 4-point scale in terms of the occurrence of the behaviour over the past six months. The labels are 0 = *never or rarely*, 1 = *sometimes*, 2 = *often*, or 3 = *very often*. An alternate way of scoring these symptom ratings is to take ratings of 0 and 1 as symptoms not being present, and ratings of 2 and 3 as symptoms being present. The internal reliability (Cronbach's alpha) value for full set of the eight ODD symptoms was .92 for the study sample. For IA and HI symptoms they were .95 and .94, respectively. The

total scores for IA and HI scales were used to examine how the factors in the optimum ODD model were related to ADHD.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). The SDQ is a rating scale for screening the emotional and behavioral problems of children and adolescents, aged 4 to 16 years (Goodman, 1997; Goodman, Meltzer, & Bailey, 1998). The study used the teacher versions of the SDQ. It has 25 items, with five scales (Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Relationship Problems, and Prosocial Behavior). Each scale has five items, and each item is rated by the informant as either "not true" (scored 0), "somewhat true" (scored 1), or "certainly true" (scored 2). The internal consistency (Cronbach's alpha) values for maternal ratings for the current sample were .76 (Emotional Symptoms), .72 (Conduct Problems), .87 (Hyperactivity/Inattention), .55 (Peer Relationship Problems), and .82 (Prosocial Behavior). The total scores for each of these scales were used to examine the external validity of the factors in the optimum ODD model.

Procedure

Stratified random sampling was used to select schools to approach for participation in the study. The population was divided into nine groups, corresponding to the nine regions of the State of Victoria, Australia. A total of 15 elementary schools from the nine regions were contacted. Within each region, a random number table was used to determine the schools to be contacted. Of the schools contacted 10 consented to participate.

Following consent from directors of education and school principals, classroom teachers were issued the appropriate number of large sealed envelopes to be forwarded to mothers, through their students. Each envelope contained a letter describing the study, the consent form, a return envelope, a form for background information, the DBRS, and the SDQ. To minimise bias

in ratings, the letter to mothers indicated that the study was addressing aspects of children's home and school behaviours, and the questionnaires were not identified by name.

Parents were requested to provide the child's age, gender, and ethnic background. The parent consent form asked parents to indicate their willingness to have the DBRS and SDQ completed by their child's class teacher. In all 61% of parents (N = 328) consented. When such consent was available, the child's teacher was requested to complete the DBRS and SDQ for the child. In all, 65% (N = 213) of the teachers who were requested to complete these questionnaires provided complete ratings for these questionnaires.

Statistical Analysis

All statistical analyses were conducted using M*plus* Version 7 (Muthén & Muthén, 2012). In the one-factor CFA model, all the ODD symptoms loaded onto one latent factor. In the two-factor and three-factor CFA models, all the relevant ODD symptoms loaded onto their own respective factors, and these factors were correlated. As the Stringaris and Goodman (2009) model has only 1 symptom for the hurtful factor, the error variance for this symptom was fixed to 0.093 (the value for the error variance for this symptom in the one-factor model) for identification purposes. For all other models, error variances were freely estimated, and the latent variances were set to unity for model identification. For the bifactor models, the loadings for the specific factors were identical to their corresponding first-order factor models. In addition, and for all bifactor models, all symptoms loaded on one general ODD factor. Also, latent factors (general and specific) were uncorrelated, and error variances were freely estimated.

To test the fit of the various CFA models, we used the mean and variance-adjusted weighted least squares (WLSMV) estimator. This is a robust estimator, ideally suited for ordered-categorical scores (Muthen & Muthen, 2010). As χ^2 values, including the WLSMVχ^2, are

inflated by large sample sizes, model fit was also evaluated by three approximate or practical fit indexes. The indexes used were the root mean squared error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker Lewis Index (TLI). The guidelines suggested by Hu and Bentler (1998) are that RMSEA values close to 0.06, or below, be taken as good fit, 0.07 to 0.08 as moderate fit, 0.08 to .10 as marginal fit, and >.10 as poor fit. For the CFI and TLI, values close to .95 or above indicate good fit, and values close to .90 and .95 are taken as acceptable fit.

To test the external validities of the factors in the optimum ODD model, the analysis for the SDQ and ADHD constructs were examined separately. In each instance, the relevant constructs were correlated with each other and with the relevant ODD factors in this optimum model. To evaluate the magnitude of the correlations, we used Cohen's (1992) effect size criteria for Pearson correlation coefficents of < .1 = negligible, .1 = small, .3 = medium, and .5 = large.

Results

Initial examination indicated no missing values in the dataset. Also, based on recoded scores (ratings of 0 and 1 = symptom not present, and ratings of 2 and 3 = symptoms present), 9 of 212 (4.24%) of the children rated had 4 (i.e., the threshold number of symptoms needed for an ODD diagnosis) or more symptoms present.

Table 1 shows the fit of all the CFA models that were tested. As shown, the analyses for the bifactor models, based on Aebi et al's (2010) and Burke et al's (2010) models, resulted in inadmissible solutions. Based on Hu and Bentler's (1998) guidelines, the one-factor model and Aebi et al.'s (2010) and Burke et al.'s three-factor (2010) models showed good fit in terms of their CFI and TLI values. The RMSEA values for these models indicated either moderate or marginal fit. In contrast, Rowe et al.'s (2010) two-factor model, Stringaris and Goodman's (2009) three-factor model, and the bifactor model based on Stringaris and Goodman's (2009)

model showed good fit in terms of their RMSEA, CFI, and TLI values. However, of these models, the chi-square value for the bifactor model based on Stringaris and Goodman's (2009) model was not significant, whereas they were significant for the other two models. Additional analyses indicated that the bifactor model based on Stringaris and Goodman's (2009) model showed better fit than Rowe et al.'s (2010) two-factor model ($\Delta df = 6$; ΔWLSMV$\chi^2 = 12.62$, $p < .05$), and no difference in fit between the bifactor model based on Stringaris and Goodman's (2009) model and the Stringaris and Goodman (2009) three-factor model, ($\Delta df = 5$; ΔWLSMV$\chi^2 = 8.81$, $p = .1169$). However, as the chi-square value for the bifactor model based on Stringaris and Goodman's (2009) model was not significant, and the chi-square value for Stringaris and Goodman's (2009) three-factor model was significant, it can be interpreted that there is more support for the bifactor model based on Stringaris and Goodman's (2009) model. Besides the general factor, this model has specific factors for irritable, headstrong, and hurtful. Further support for a general factor is indicated by the very high intercorrelations among the factors in Stringaris and Goodman's (2009) model. They were .91 between irritable and headstrong, .82 between irritable and headstrong, and .80 between headstrong and hurtful. Thus, the bifactor model based on Stringaris and Goodman's (2009) model was taken as the optimum ODD model.

Table 2 shows the factor loadings for the factors in the bifactor model based on Stringaris and Goodman's (2009) model. As shown, all eight ODD symptoms had salient loadings (loadings $\geq .40$; Field, 2013) on the general factor. In contrast, for the specific factors, only the loadings for touchy on the irritable factor, and spiteful/vindictive on the hurtful factor were salient. Table 3 includes the omega hierarchical (ω_h), and the explained common variance (ECV). The ECV for the general factor was .86, and the ECV for irritable, headstrong, and hurtful specific factors were .06, .03, and .05, respectively. The ω_h for the general factor was .95, and the

ω_h values for irritable, headstrong, and hurtful specific factors were .15, .00, and .33, respectively. Thus, based on the ECV and ω_h values for the bifactor model based on Stringaris and Goodman's (2009) model, only the general factor in this model can be meaningfully interpreted.

Table 3 shows the correlation coefficients for the SDQ constructs with the factors in the bifactor model based on Stringaris and Goodman's (2009) model. As shown, the general factor correlated with all the SDQ constructs. With the exception of prosocial behaviour, all correlations were positive. The correlation involving prosocial behaviour was negative. The specific factors for headstrong and hurtful did not correlate significantly with any of the SDQ constructs. The specific factor for irritable correlated positively with emotional symptoms and peer relationship problems. While the correlation for irritable with emotional symptoms, and peer relationship problems were large and medium respectively, the correlation for the general factor with emotional symptoms, and peer relationship problems were both small. Table 3 also includes the correlation coefficients for the ADHD constructs with the factors in the bifactor model based on Stringaris and Goodman's (2009) model. The table shows that both IA and HI were correlated positively with the specific factor for headstrong and the general factor, and no associations were observed between irritable and hurtful factors and IA and HI. Both the significant associations involving the general factor were large, whereas both the significant associations involving the headstrong factor were small.

Discussion

The major aim of the study was to establish the optimum structural model for DSM-IV/DSM-5 ODD symptoms. The study used CFA to examine one-factor (inspired by DSM-IV), two-factor (Rowe et al., 2010), three-factor (Aebi et al., 2010; Burke et al., 2010; Stringaris &

Goodman, 2009) models, bifactor versions of all these models, and the bifactor model proposed by Burke et al. (2014). The findings indicated some support for all the first-order models tested, with more support for Rowe et al.'s (2010) two-factor model and Stringaris and Goodman's (2009) three-factor model. With the exception of the bifactor models based on Stringaris and Goodman's (2009) model, the analyses for all the other bifactor models resulted in inadmissible solutions. The bifactor models based on Stringaris and Goodman's (2009) model showed better fit than Rowe et al.'s (2010) two-factor model. Although the bifactor model based on Stringaris and Goodman's (2009) model showed comparable fit with Stringaris and Goodman's (2009) model, the chi-square value for the bifactor model based on Stringaris and Goodman's (2009) model was not significant, whereas it was significant for Stringaris and Goodman's (2009) model. Additionally, there were very high intercorrelations among the factors (ranging from .89 to .91) in the Stringaris and Goodman's (2009) model. Given these findings, the bifactor model based on Stringaris and Goodman's (2009) model was interpreted as the optimum ODD model.

The findings supporting the different first-order factor one-, two-, and three-factor model were as predicted, and are, in part, consistent with existing data (e.g., Aebi et al., 2010; Burke et al., 2005; Burke et al., 2010; Krieger et al., 2013; Rowe et al., 2010; Stringaris & Goodman, 2009). For example, while Krieger et al. (2013) found better fit for Stringaris and Goodman's (2009) three-factor model over other proposed models, Herzhoff and Tackett (2016) reported better fit for Burke et al.'s (2005) two-factor model over other models, including Stringaris and Goodman's (2009) model. Although, based on Burke et al.'s (2014) study, we expected some support for Burke et al.'s (2014) bifactor model, the analysis for this model failed to provide an admissible solution. However, our interpretation of most support for the bifactor model based on Stringaris and Goodman's (2009) model concurs with our prediction that there will be relatively

more support for one or more bifactor models. Although some of our findings are comparable to existing findings, they also extend existing findings as this is the first study to demonstrate the applicability of a bifactor ODD model in primary school children, based on teacher reports.

A second aim of the study was to examine the external validity of the factors in our optimum model. In this respect, the general factor correlated (in the theoretically expected direction), and had large effect sizes, with conduct problems, hyperactivity/inattention and prosocial behaviour. Further, the general factor correlated, and had small effect sizes, with emotional symptom and peer relationship problems. The specific factor for irritable correlated (with a large effect size) with emotional symptoms, and peer relationship problems (medium effect size). While the correlations for Irritable with emotional symptoms, and peer relationship problems were large and medium, respectively. Neither the specific factors for headstrong or hurtful correlated significantly with any of the SDQ constructs. While IA and HI correlated (large effects sizes) with the general factor, they also correlated with headstrong (small effect sizes). The specific factors for irritable and hurtful were not associated with IA and HI. Taken together, the findings involving the SDQ and ADHD constructs can be interpreted as supporting the external validity of all the factors in the bifactor model based on Stringaris and Goodman's (2009) model.

The findings showed that the optimum bifactor model, based on Stringaris and Goodman's (2009) model, had salient loadings (loadings $\geq .40$; Field, 2013) for all symptoms on the general factors. In contrast, for the specific factors, only the loadings for touchy on the irritable factor, and spiteful/vindictive on the hurtful factor were salient. Additionally, all symptoms had higher loadings on the general factor than the specific factor for which they served as indictors. None of the symptoms had significant loadings for the headstrong specific factor.

The ECV for the general factor was .86, and the ECV for Irritable, Headstrong, and Hurtful specific factors were .06, .03, and .05, respectively. Also, the ω_h for the general factor was .95, and the ω_h values for irritable, headstrong, and hurtful specific factors were .15, .00, and .33, respectively. Since ω_h value can be interpreted as an index of internal consistency reliability, these findings indicate good reliability for the general factor. Furthermore, once the variances due to the general factor are removed, the specific factors have extremely low reliabilities. According to Reise et al. (2013), high ECV and ω_h values of at least 0.75 are preferred for meaningful interpretation of a scale. Based on these guidelines, the findings can be taken to mean that, although the bifactor model based on Stringaris and Goodman's (2009) three-factor model could be the optimum model and the factors have external validity, only the general factor has sufficient reliability for meaningful interpretation. Thus, although the total score confounds general and specific factor variances, it is highly saturated with general factor variances. Therefore, total scores can be used with a sufficient level of confidence.

In relation to the specific factors, as the hurtful factor has only one item and this item had salient loadings on this factor, it can be argued that the hurtful factor can be meaningfully interpreted. In contrast, as none of the symptoms had significant loadings for the headstrong specific factor, these symptoms have no common variance over and above what is shared with the general ODD factor. Thus, when applying the bifactor model based on Stringaris and Goodman's (2009) model, it is important to keep in mind that the headstrong factor may be of little consequence to the model, as also supported by the extremely low EVC value, and zero ω_h value for this factor. Although the irritable factor had low EVC value, and low ω_h value, it showed a high correlation with SDQ emotional symptoms, thereby suggested that the specific factor for irritable could be valuable. Taken together, these interpretations suggest that a modified

bifactor model without a specific factor for headstrong may be viable (i.e., a general factors on which all 8 symptoms load and specific factors for only irritable and hurtful). Regardless, and as noted earlier, the specific factors in a bifactor model are not the same as the first-order factors found in the corresponding first-order factor model. The specific factors represent common variance after controlling for the general ODD factor, whereas the first-order factors also include variances that are part of the general ODD factor.

The findings in the study have taxonomic, diagnostic, practical, and research implications. Taxonomically, our findings align closely with how the ODD symptoms are presented in DSM-5. As noted previously, in the DSM-5, the groupings and item content within these groups were based on the three-factor model proposed by Stringaris and Goodman's (2009). Given the support found in the study for Stringaris and Goodman's (2009) three-factor model, it can be argued that the findings concur with how the ODD symptoms are grouped in the DSM-5. Diagnostically, when applied to the DSM-5, our support for a meaningful interpretation for only the general ODD factor means that, despite the possibility that the DSM-5 ODD symptoms are multidimensional, they could be grouped and considered together. This suggests that the current practice of diagnosing ODD, as presented in DSM-5 (in terms of all symptoms being considered together), is appropriate. Practically, although the findings and interpretations made in the study are directly relevant to the DBRS (Barkley & Murphy, 1998), it can be speculated that they could be relevant to other ODD rating scales and ODD in general. Additionally, as the DSM-5 views severity of ODD in terms of the presence of ODD across settings, the use of teacher reports as sources of information for clinical diagnosis seems warranted. The findings here are, to a certain extent, comparable with previously published parental ratings of ODD, and thus the evaluation of severity of ODD in terms of parent and teacher reports would be meaningful. From a research

viewpoint, the low variances and internal consistency reliabilities of the group factors raises the possibility that past findings for these factors may be of little value. Relatedly, if we wish to obtain unbiased estimates of the relationships of external correlates with the general and the specific factors, CFA (as used in the current study) is recommended. This method allows researchers to ascertain the true meaning of these "purified" factors.

The findings and interpretations made in this study should be considered with a few limitations in mind. First, this study obtained scores for the ODD symptoms using the DBRS (Barkley & Murphy, 1998), a rating scale. Thus, it is uncertain if different results would emerge with other rating scales or interview-based data. Unlike rating scales, clinical interviews provide opportunities for clinicians to deal with respondents' uncertainties when answering questions. Second, it is conceivable that ratings of ODD symptoms are influenced by a number of background factors (e.g., age, socio-economic status, and comorbidity). As this study did not control for these factors, the findings may be confounded by them. Third, this study used teacher reports and, as such, it is uncertain if similar findings would emerge with parent reports, or with self-reports. Fourth, all the participants in this study were from the general school community, based in a Western country. Thus, it is possible that the findings may not be applicable to clinic-referred samples, samples of children with a diagnosis of ODD, or to different cultural and national groups. Fifth, it is conceivable that low sample size (N = 213) may have contributed to non-admissible solutions for some of the models tested. As such, it cannot be ruled out that, with larger samples, some of these models, including the bifactor models, may be better supported than the three-factor model proposed by Stringaris and Goodman's (2009). Given these limitations, some may wish to consider the results of this study as being preliminary. Although more studies in this area are required, it is argued that the results from this study provide

sufficient support for the view that future studies, controlling for the limitations mentioned, could benefit from applying the CFA approach to help resolve questions regarding the factor structure of ODD symptoms.

References

Aebi, M., Muller, U.C., Asherson, P., Banaschewski, T., Buitelaar, J., Ebstein, R., Eisenberg, J., Gill, M., Manor, I., Miranda, A., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J., Sonuga-Barke, E., Thompson, M., Taylor, E., Faraone, S.V., & Steinhausen, H.C. (2010). Predictability of oppositional defiant disorder and symptom dimensions in children and adolescents with ADHD combined type. *Psychological Medicine*, *40*(12), 2089-2100.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*.Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (Text Revision).* Washington, DC: Author.American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.).* Washington, DC: Author.

Australian Bureau of Statistics [ABS]. (2007). *Population census*. Canberra: Australian Government Publishing Service.

Australian Bureau of Statistics. (1997). *Australian standard classification of occupations*. Canberra: Australian Government Publishing Service.

Barkley, R. A., & Murphy, K. R. (1998). *A clinical workbook: Attention-deficit hyperactivity disorder*. New York: Guilford.

Burke, J. D. (2012). An affective dimension within oppositional defiant disorder symptoms among boys: personality and psychopathology outcomes into early adulthood. *Journal of Child Psychology and Psychiatry*, *53*(11), 1176-1183.

Burke, J. D., Boylan, K., Rowe, R., Duku, E., Stepp, S. D., Hipwell, A. E., & Waldman, I. D. (2014). Identifying the irritability dimension of ODD: Application of a modified bifactor model across five large community samples of children. *Journal of Abnormal Psychology*, *123*(4), 841-851.

Burke, J. D., Loeber, R., Lahey, B. B., & Rathouz, P. J. (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. *Journal of Child Psychology and Psychiatry, 46*, 1200–1210.

Burke, J. D., Waldman, I., & Lahey, B. B. (2010). Predictive validity of childhood oppositional defiant disorder and conduct disorder: Implications for the DSM-V. *Journal of Abnormal Psychology*, *119*(4), 739-751.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796-846.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Ezpeleta, L., Granero, R., de la Osa,Penelo, N., & Domenech, J.M. (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. *The Journal of Child Psychology and Psychiatry, 53*(11), 1128-1138.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th revised edition). London: Sage Publications Ltd.

Gomez, R., (2016). Factor Structure of Parent and Teacher Ratings of the ODD Symptoms for Malaysian Primary School Children *Asian Journal of Psychiatry*.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *The Journal of Child Psychology and Psychiatry, 38*(5), 581-586.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, *7*(3), 125-130.

Herzhoff, K., & Tackett, J. L. (2016). Subfactors of oppositional defiant disorder: Converging evidence from structural and latent class analyses. *Journal of Child Psychology and Psychiatry*, *57*(1), 18-29.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424.

Krieger, F. V., Polanczyk, G. V., Goodman, R., Rohde, L. A., Graeff-Martins, A. S., Salum, G., Gadelha, A., Pan, P., Stahl, D., & Stringaris, A. (2013). Dimensions of oppositionality in a Brazilian community sample: Testing the DSM-5 proposal and etiological links. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(4), 389-400.

Lavigne, J. V., Bryant, F. B., Hopkins, J., & Gouze, K. R. (2015). Dimensions of oppositional defiant disorder in young children: Model comparisons, gender and longitudinal invariance. *Journal of Abnormal Child Psychology*, *43*(3), 423-439.

McDonald, R. P. (1999). *Test theory: A unified approach*. New Jersey: Lawrence Erlbaum Associates.

Muthén, L. K., & Muthén, B. O. (2012). Mplus statistical modeling software: Release 7.0. *Los Angeles, CA: Muthén & Muthén*.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129-140.

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*(4), 726.

Stringaris, A., & Goodman, R. (2009). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, *50*(3), 216-223.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133.

Table 1

Fit of the Factor Models to the ODD Symptom Ratings

Model	WLSMV χ^2	*df*	RMSEA(90% CI)	CFI	TLI
One-, Two- and Three-Factor Models					
One-factor	52.99**	20	.088 (.060- .117)	.99	.99
2-factor (Rowe et al., 2010)	32.63*	19	.058 (.020 - .091)	.99	.99
3-factor (Aebi et al., 2010)	34.17**	17	.069 (.034 - .102)	.99	.99
3-factor (Burke et al., 2010)	35.53**	17	.072 (.038 - .105)	.98	.98
3-factor (Stringaris & Goodman, 2009)	31.08*	18	.058 (.020 - .091)	.99	.99
Bifactor Models					
2-specific factors (Rowe et al., 2010)		Non admisable solution			
3-specific factors (Aebi et al., 2010)		Non admisable solution			
3-specific factors (Burke et al., 2010)		Non admisable solution			
3-specific factors (Stringaris & Goodman, 2009)	22.30	13	.068 (.000 - .098)	.99	.99
2-specific factors (Burke et al., 2014)		Non admisable solution			

Note. CFI = comparative fit index; CI = confidence interval; RMSEA = root mean square error of approximation; TLI = Tucker Lewis Index. As the Stringaris and Goodman (2009) has only 1

symptom for one its factors, the error variance for this symptom was fixed to 0.093 (the value for the error variance for this symptom in the one-factor model) for identification purposes.

$^{**}p < .01, ^{*}p < .05$.

Table 2

Factor Loadings for the Items in the Bifactor Model with Specific Factors Based on Stringaris and Goodman's (2009) *3-Factor Model*

Brief item description	Mean	SD	Irritable	Headstrong	Hurtful	General
Temper (1)	.23	.55	-.34			.84
Touchy (6)	.46	.72	-.40			.79
Angry (7)	.26	.61	-.38			.87
Argues (2)	.26	.56		-.23		.95
Defies (3)	.18	.50		-.14		.92
Annoys (4)	.35	.65		.06		.89
Blames (5)	.42	.70		.39		.92
Spiteful/vindictive (8)	.13	.39			.57	.80
Omega (ω)			.94	.97	.97	.98
Omega hierarchical (ω_h)			.15	.00	.33	.95
Variance Explained						
Total variance			,05	.03	.04	77
Common variance			.06	.03	.05	.86

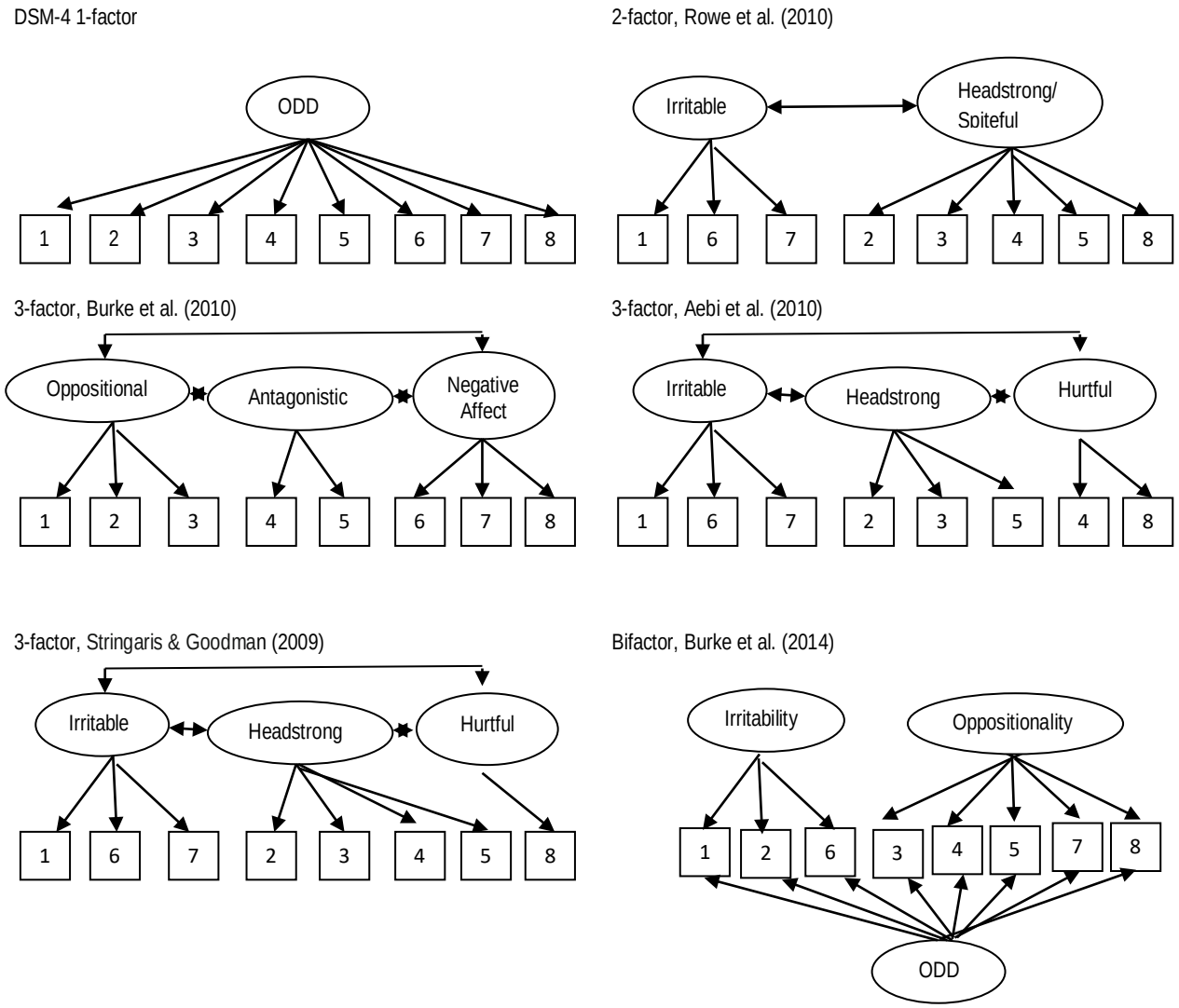
Table 3

Correlation Coefficients of SDQ and ADHD Constructs with the Factors in the Bifactor Model with Specific Factors Based on Stringaris and Goodman's (2009) 3-Factor Model

	Irritable	Headstrong	Hurtful	General
SDQ constructs				
Emotional Symptoms	.60***	-.02	.02	.21**
Conduct Problems	.19*	.07	-.01	.63***
Hyperactivity/Inattention	-.33	.09	-.13	.67***
Peer Relationship Problems	.35**	-.02	-.03	.28***
Prosocial Behaviour	-.07	-.14	-.20	-.45***
ADHD constructs				
Inattention	-0.04	0.24*	-0.01	0.58***
Hyperactivity/Impulsivity	0.09	0.17*	-0.16	0.66***

Note. To evaluate the magnitude of the associations, we used Cohen's (1992) effect size criteria for Pearson correlation coefficients of < .1 = negligible.1 = small, .3 = medium and .5 = large.

***p < .01, **p < .0*p < .05.



Note. 1 = often loses temper; 2 = often argues with authority figures or with adults; 3 = often actively defies or refuses to comply with requests from authority figures or with rules; 4 = often deliberately annoys others; 5 = often blames others for his or her mistakes or misbehavior; 6 = is often touchy or easily annoyed; 7 = is often angry and resentful; 8 = has been spiteful or vindictive at least twice within the past 6 months.

Figure 1. Common currently proposed Oppositional Defiant Disorder models.