Article:
Burke, J.D., Boylan, K., Rowe, R. et al. (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). pp. 841-851. ISSN: 0021-843X

https://doi.org/10.1037/a0037898

Identifying the Irritability Dimension of ODD: Application of a Modified Bifactor Model Across Five Large Community Samples of Children

Jeffrey D. Burke, Ph.D.
University of Pittsburgh

Khrista Boylan, Ph.D., M.D.
McMaster University

Richard Rowe, Ph.D.
University of Sheffield

Eric Duku, Ph.D.
McMaster University

Stephanie D. Stepp, Ph.D.
University of Pittsburgh

Alison E. Hipwell, Ph.D.
University of Pittsburgh

Irwin D. Waldman, Ph.D.
Emory University

Author Note

Jeffrey D. Burke, Department of Psychiatry, University of Pittsburgh; Khrista Boylan, Department of Psychiatry, McMaster University; Richard Rowe, Department of Psychology, University of Sheffield; Eric Duku, Department of Psychiatry, McMaster University; Stephanie D. Stepp, Department of Psychiatry, University of Pittsburgh; Alison E. Hipwell, Department of Psychiatry, University of Pittsburgh; Irwin D. Waldman, Department of Psychology, Emory University.

This work was supported by a grant to Dr. Burke (MH095969) from the National Institute of Mental Health (NIMH), a grant to Dr. Benjamin B. Lahey (R01 MH59111) from NIMH, grants to Dr. Rolf Loeber from the National Institute of Mental Health (Grant 2013-JF-

FX-0058) and from the National Institute of Drug Abuse (Grant 2RO1DA012237-11). The UK Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provide core support for ALSPAC. The authors are grateful to all the families who took part in the Avon Longitudinal Study of Parents and Children (ALSPAC) study and to the ALSPAC team. This publication is the work of the authors and Drs. Burke, Boylan, Rowe, Duku, Hipwell, Stepp and Waldman will serve as guarantors for the contents of this paper.

Correspondence should be directed to Jeffrey D. Burke, Ph.D., Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213; email: burkejd@upmc.edu.

Abstract

The importance of irritability as measured among the symptoms of oppositional defiant disorder (ODD) has dramatically come to the fore in recent years. New diagnostic categories rely on the distinct clinical utility of irritability, and models of psychopathology suggest it plays a key role in explaining developmental pathways within and between disorders into adulthood. However, only a few studies have tested multidimensional models of ODD, and the results have been conflicting. Further, consensus has not been reached regarding which symptoms best identify irritability. The present analyses use data from five large community data sets with five different measures of parent-reported ODD, comprising 16,280 youth in total, to help resolve these questions. Across the samples, ages ranged from 5 to 18, and included both boys and girls. Confirmatory factor analyses demonstrated that a modified bifactor model showed the best fit in each of the five data sets. The structure of the model included two correlated specific factors (irritability and oppositional behavior) in addition to a general ODD factor. In four of the five models, the best fit was obtained using the items of being touchy, angry and often losing temper as indicators of irritability. Given the structure of the models and the generally high correlation between the specific dimensions, the results suggest that irritability may not be sufficiently distinct from oppositional behavior to support an entirely independent diagnosis. Rather, irritability may be better understood as a dimension of psychopathology that can be distinguished within ODD, and which may be related to particular forms of psychopathology apart from ODD.

Evidence illustrating the link between oppositional defiant disorder (ODD) and depression (Boylan, Vaillancourt, Boyle, & Szatmari, 2007; Burke, Loeber, Lahey, & Rathouz, 2005; Copeland, Shanahan, Costello, & Angold, 2009; Rowe, Maughan, & Eley, 2006) generated several hypotheses pertaining to the identification and the developmental course of mood disorders. Researchers have demonstrated that irritability as a distinct dimension within ODD symptoms robustly predicts depression and anxiety (Burke & Loeber, 2010; Burke, 2012; Burke, Hipwell, & Loeber, 2010; Hipwell et al., 2011; Rowe, Costello, Angold, Copeland, & Maughan, 2010; Stringaris, Cohen, Pine, & Leibenluft, 2009; Stringaris & Goodman, 2009a, 2009b). Studies using latent classification techniques show that irritability symptoms may distinguish groups of children (Burke, 2012; Kuny et al., 2013), and that those with irritability features are at greater risk for depression and anxiety in adulthood.

This relatively recent evidence for irritability as one of multiple dimensions of ODD stands in some contrast to the unidimensional model of ODD symptoms evident in the existing literature. Factor analytic studies (e.g. Burns, Boe, Walsh, Sommers-Flanagan, & Teegarden, 2001; Hartman et al., 2001; Pelham, Gnagy, Greenslade, & Milich, 1992) of DSM symptoms have tended to find that ODD symptoms hang together relative to other disorders. A meta-analysis focused on ODD and CD symptoms alone (Frick et al., 1993) found that the symptoms of the two disorders aligned along two dimensions. Notably, however, the symptoms of ODD largely clustered together in one quadrant, distinct from CD. In a recent paper using a hierarchical approach to examine the structure of the symptoms of the disruptive behavior disorders, Bezdjian and colleagues (Bezdjian et al., 2011) found evidence that a general externalizing factor explained approximately one-quarter of the total variance, and illustrated the specificity of ODD symptoms relative to ADHD and CD. Further, whereas the symptoms of

each of ADHD and CD broke across multiple components, ODD symptoms did not, showing strong evidence of unidimensionality (Bezdjian, et al., 2011). In summary, a range of existing evidence would broadly support a conclusion that the symptoms of ODD are distinct from other disorders and that they represent an underlying unidimensional construct.

Given this evidence for unidimensionality, recent findings of multiple meaningful dimensions among ODD symptoms (e.g. Burke, 2012; Rowe, et al., 2010; Stringaris & Goodman, 2009b) present a bit of a puzzle. If the bulk of the literature shows evidence for unidimensionality among ODD symptoms, does this call into question the validity of evidence for multidimensionality? It is important to note that the evidence for unidimensionality within ODD symptoms comes exclusively, to our knowledge, from studies of ODD in the context of other disorders, which may influence the degree to which multidimensionality within ODD is apparent. When ODD symptoms, independent from the symptoms of other disorders, are examined via factor analysis, distinct dimensions are evident (Aebi et al., 2010; Burke, et al., 2010; Ezpeleta, Granero, de la Osa, Penelo, & Domenech, 2012; (Krieger et al., 2013), Rowe, et al., 2010).

There has been conflicting evidence as to how many dimensions best characterize the structure within ODD symptoms. Stringaris & Goodman (2009b) proposed three dimensions on a priori grounds- Headstrong (including arguing with adults, purposefully annoying others, disobedience and blaming others for own mistakes) Irritable (including temper tantrums, anger and touchiness) and Hurtful (including spite and vindictiveness). Stringaris and Goodman did not present factor analyses to support their conceptualization in their original work, but demonstrated a range of specific associations between the dimensions and other disorders in cross-sectional and longitudinal analyses. For example, the irritable dimension was specifically associated with

emotional disorders, headstrong was associated with ADHD and hurtful was associated with aggressive CD symptoms and callous-unemotional traits. Support for this Headstrong, Irritable and Hurtful model has subsequently been provided in a sample of children aged 5-17 years with ADHD (Aebi et al. 2010) and in a Brazilian sample aged 6-12 years (Krieger et al, 2013). The latter study replicated the specific associations of the dimensions with concurrent psychopathology and also provided evidence of etiological specificity. For example irritability was associated with a family history of depression. Substantial correlations between the factors (irritable-headstrong, r=.73, irritable-hurtful r=.53, headstrong-hurtful, r=.59) were also reported. High correlations between the irritability and behavioral dimensions of ODD symptoms have been reported elsewhere (e.g. r=.55 (Rowe, et al., 2010) and r=.89 (Aebi, et al., 2010). The Headstrong, Irritable and Hurtful model has been adopted within the DSM 5 definition of ODD (American Psychiatric Association, 2013).

However, evaluation of the underlying structure of ODD symptoms has been far from exhaustive and where tests have been conducted, the Irritable-Hurtful-Headstrong model has not been consistently supported. Only three studies testing models of multiple ODD dimensions have included formal tests of a unidimensional model, and they have produced conflicting findings. Krieger et al. (2013) and Aebi and colleagues (Aebi, et al., 2010) found evidence of poor fit for the unidimensional model of ODD. On the other hand, Ezpeleta and colleagues (Ezpeleta, et al., 2012) found no statistical justification to accept a multidimensional model of ODD over a single dimension among preschool children oversampled for behavioral problems.

Furthermore, factor analytic models including the hurtful dimension are not in fact testable when measures follow the DSM approach of treating spite and vindictiveness as a single item (e.g. Ezpeleta, et al., 2012, Rowe et al, 2010). Recent evidence further calls into question

the utility of the Hurtful dimension, in that it did not predict any tested outcomes in a large community sample, whereas distinct predictions were observed from the Headstrong factor to conduct problems, and from the Irritable factor to depression (Whelan, Stringaris, Maughan, & Barker, 2013).

In addition, evidence on the symptoms that make up the dimensions has been inconsistent. Burke and colleagues, in a community sample of girls (Burke et al., 2010) and a clinical sample of boys (Burke, 2010) identified a negative affect factor which had much similarity to the irritable dimension of the Irritable-Hurtful-Headstrong model (Stringaris & Goodman, 2009a). However, it was indexed by a slightly different symptom set (anger, touchiness and spitefulness). The three factors described by Burke et al. (2010) also included an oppositional factor including temper tantrums, arguing and defiance and an antagonistic factor including annoying and blaming others (Burke, 2012; Burke, et al., 2010). Three studies have explicitly compared these two conceptualizations, with mixed results. Krieger et al (2013) found a clear preference for the Stringaris & Goodman (2009b) approach. On the other hand, Lavigne, Gouze, Bryant & Hopkins (2014) found greater support for the negative affective model of irritability (e.g. Burke & Loeber, 2010). Finally, Ezpeleta, et al. (2012) found support for both of the aforementioned models of irritability. The differences between these negative affect and irritability constructs are arguably overshadowed by their consistency: each features anger and irritability, and both predict problems with depression and anxiety (Burke, et al., 2010; Rowe, et al., 2010; Stringaris & Goodman, 2009a). The inconsistencies that do exist nevertheless create a number of difficulties. Without a consistent model, researchers are challenged to efficiently examine the issue without having to either examine multiple models or to opt for one particular model over another arbitrarily. In addition, given the aforementioned establishment of a

particular structure for the ODD items in the DSM 5 (American Psychiatric Association, 2013), and given the development of a new measure of irritability based on the emerging literature (Stringaris et al., 2012), resolving discrepancies in the identification of irritability has clear applied utility as well.

The investigation of dimensionality. A general increase in the investigation of dimensionality underlying currently established constructs has provided examples of the use of multidimensional modeling to address questions such as those raised by the current literature on irritability within ODD. Although arguments for dimensional conceptualizations have a long history, a marked shift in focus was particularly evident in the development towards the DSM 5 (Helzer, Kraemer, & Krueger, 2009; Moffitt et al., 2008). Concerns that disorders are often heterogeneous and demonstrate high levels of comorbidity with multiple other disorders supported efforts to consider not only multidimensionality, but also models of latent structure that extend beyond simple, hierarchical correlated factors model.

One particular modeling strategy, bifactor modeling has been used to model specific sub-dimensions of disorder in the context of a general overall dimension in a number of symptom domains. In bifactor models, a general factor reflects the common variance among all items within a construct, while specific factors reflect additional common variance among distinct groups of items. It is well-suited for representing the "construct-relevant multidimensionality" arising when a broad construct includes conceptually narrow subdomain constructs (Reise, 2012). For example, the approach has been used to confirm that ADHD consists of a single broad dimension along with separate and distinct inattention and hyperactivity-impulsivity dimensions (Martel, von Eye, & Nigg, 2010). Martel and colleagues (2010) note that a bifactor model of ADHD has several important implications: the potential for distinct etiological inputs,

distinct treatment needs for individuals with different constellations of inattentive versus hyperactivity-impulsivity symptoms, and potentially different outcomes arising from the separate symptom dimensions.

A number of other examples illustrate the utility of bifactor modeling for testing multidimensionality within a construct. It has been used to confirm expected distinctions among multiple diagnostic categories (e.g. major depression, mania, bulemia) within a single self-report measure (Gibbons, Rush, & Immekus, 2009), or a general component of disruptive behavior along with specific components of ADHD and ODD (Martel, Gremillion, Roberts, von Eye, & Nigg, 2010). It has provided evidence for cross-cutting dimensional commonalities, as seen in models of with a common negative affective dimension along with distinct narrow anxiety and depression dimensions (Bados, Gomez-Benito, & Balaguer, 2010; Simms, Gros, Watson, & O'Hara, 2008). Bifactor modeling has also been used to confirm a model that includes a general psychopathy factor and specific subdimensions on the Psychopathy Checklist – Revised (Patrick, Hicks, Nichol, & Krueger, 2007).

In the context of current issues regarding irritability, the bifactor modeling strategy offers the potential for generating compelling evidence. It is particularly relevant, given that questions remain about whether or not the structure of ODD involves a general dimension or two distinct dimensions (or both) and how putatively distinct dimensions might relate to one another. It may shed light on distinct etiological targets, intervention targets or differential risks, as noted by Martel and colleagues (Martel, von Eye, & Nigg, 2010) regarding their examination of ADHD. The results may have particular significance regarding current diagnostic controversies regarding irritability. Specifically, if the data do not support a general ODD factor, or if there is evidence in favor of an orthogonal specific dimension of irritability (relative to a behavioral

factor) this would support the argument that a diagnostic construct, such as DMDD, could be usefully separated from ODD.

The present study uses confirmatory factor analyses in five large existing data sets to test competing models of irritability within ODD symptoms. Single and multidimensional models, including the bifactor model, will be tested. Elements of the measurement and structural models will be examined, competing models of irritability will be tested, and the magnitude of the relations between specific dimensions will be identified. The specific questions to be addressed with these analyses are: 1) Do the data support distinguishing a dimension of irritability within ODD? 2) Is there evidence, replicated across data sets, to support one of the proposed models of irritability among ODD symptoms? 3) Is there evidence for a general dimension of ODD in addition to specific dimensions?

Method

Studies and Measures. The present study used existing data from five large community samples, each with differing measures of parent-reported ODD symptoms. Informed consent was obtained for all participants in each study. Table 1 provides an overview of the sample sizes and demographics.

Pittsburgh Youth Study (PYS). The PYS is a longitudinal study of urban boys who were recruited from a random sample of public school students in grades 1, 4 and 7. The screening sample included 2,573 boys (84.7% of the total number contacted for participation). From each grade, the 30% of those with the highest rates of antisocial behavior (approximately 250) along with 250 randomly drawn from the remainder were then selected to participate in follow-up assessments. This resulted in a sample of 1,517 boys enrolled into the cohort. The data from the

first assessment wave for the selected sample were used for the present analyses, when the boys in the youngest cohort had a mean age of 7, those in the middle cohort had a mean age of 10, and those in the oldest had a mean age of 13. Approximately half of the sample was African American and half Caucasian.

The ODD symptom data used in these analyses were generated using parent report on the Revised Diagnostic Interview Schedule for Children, Parent Version (DISC-P; Costello, Edelbrock, Dulcan, Kalas, & Klaric, 1987). Parents responded as to whether or not the child had exhibited each symptom during the past six months. Thus, dichotomous values representing the presence or absence of each ODD symptom were used for these analyses.

Pittsburgh Girls Study (PGS). The PGS is a longitudinal study of a community sample of girls recruited following the enumeration of 103,238 households in the city of Pittsburgh in 1999-2000. Poorer city neighborhoods were oversampled (see Hipwell et al., 2002; Keenan et al., 2010). The present analyses use data from the first assessment wave, which included 588 five year olds, 630 six year olds, 611 seven year olds, and 622 eight year olds (total N = 2,451).

ODD symptoms were measured using the Child Symptom Inventory – 4 (CSI-4; Gadow & Sprafkin, 1994)). The CSI-4 is a standardized behavioral rating scale which maps directly onto DSM-IV symptoms, and yields both symptom counts and severity scores. Items are coded using a four-point rating scale for each item, from "never" to "very often." The measure has good reliability, including internal consistencies of .90 for ODD and test-retest reliability for symptom counts of .78 for ODD (Gadow & Sprafkin, 1994). The analyses were conducted on the ordinal, four-point parent ratings for each item.

Tennessee Twins Study (TTS). The TTS sample is representative of all 6-17 year-old twins who were born in Tennessee and living in one of the state's five metropolitan statistical

areas in 2001 (Lahey et al., 2008). These include the 28 urban, suburban, and rural counties surrounding Nashville, Memphis, Knoxville, Chattanooga, and Bristol. A random sample of families with twins was selected stratified on the age of the twins, proportional to the population of 35 geographic sub-areas. This resulted in 2,023 twin pairs (4,046 youth). The TTS consists of approximately equal numbers of monozygotic (MZ; n = 752 pairs; 51% female pairs), same-sex dizygotic (DZ; n = 670 pairs; 51% female pairs), and mixed-sex DZ pairs (n = 601 pairs). The sample was representative of the racial demographics of Tennessee in 2001.

Adult caretakers and youth were interviewed separately using the Child and Adolescent Psychopathology Scale (CAPS; Lahey et al., 2008). The CAPS covers all DSM-IV and ICD-10 symptoms of ODD, as well as ADHD, CD, major depression, and most anxiety disorders. Item administration is not grouped by disorder. Instead, the items are randomly administered, avoiding the potential for response sets, or the influence on the informant of being queried on similar content over consecutive items. Each ODD item was scored using a four point scale with value labels ranging from "not at all" to "just a little", "pretty much" and "very much" for each item. To be consistent with the eight symptoms of ODD as indicated by the DSM-IV, and with the analyses of the other data sets, the CAPS items of "being defiant" and "refusing to comply" were combined in an either/or fashion, as were the items of "being mean" and "getting even with others."

The Avon Longitudinal Study of Parents and Children (ALSPAC). The ALSPAC sample is a population based birth cohort recruited from all pregnant women resident in Bristol, UK with an expected date of delivery of April 1, 1991 to December 31, 1992. The initial ALSPAC sample consisted of 14,541 pregnancies and 13,988 children alive at one year of age. When the children were 7 years of age, the sample was bolstered with eligible cases who failed to join the

study originally, thus the available sample is greater than 14, 541. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee.

For this analysis, the cohort was aged 10 or 11 years and subjects were included if they completed a DAWBA ODD section at that time. The sample consisted of 7,420 singletons (boys = 3,740, girls=3,680). No exclusion criteria were applied.

The ALSPAC data set employs the Developmental and Well-Being Assessment (DAWBA; Goodman et al., 2000) to measure indicators of psychopathology, including the symptoms of ODD. DAWBA is an extensively evaluated structured interview enabling DSM-IV psychiatric diagnoses using items related to diagnostic criteria. Parent report on 8 DSM-IV ODD items was used for the study. The three-point item response options for the severity of each ODD item were "not more than others", "a little more than others" or "a lot more than others." The reliability for the 8 ODD items was for 0.92 for boys and 0.90 for girls. The study website contains details of all the data that is available through a fully searchable data dictionary at http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/.

Georgia Twin Study (GTS). The twin sample of the GTS comprised 846 twin pairs from the Georgia Twin Registry, a population-based registry of twins (*Mean age* = 10.60 years, *SD* = 3.20 years, age range= 6-18 years), with 49% males, 82% European Americans, 11% African Americans, 1% Hispanic Americans, and 6% mixed/other ethnicity. The sample comprised 392 (46%) monozygotic (MZ) and 454 (54%) dizygotic (DZ) twin pairs. Twins were recruited using the following procedures. In 1992 to 1993, 5,620 parents of twins born between 1980 and 1991 in the state of Georgia were contacted via mail according to the state birth records. Of these families, 1,567 twin families joined the Georgia Twin Registry, among which 846 families

provided complete ratings on the ODD items.

Symptom ratings were obtained from mothers and fathers (when the mother was not available) using the Emory Diagnostic Rating Scale (EDRS; Waldman et al., 1998). The EDRS was developed to assess symptoms of the major DSM–IV childhood psychiatric disorders. Parents rated ODD symptoms on a 0–4 scale and a symptom scale based on these items demonstrated high internal consistency in the current sample (α = .91), and the EDRS has been shown to yield ODD diagnostic rates similar to the population prevalence (Waldman et al., 1998)

Analyses. We employed Confirmatory Factor Analysis (CFA) to systematically test a set of uni- versus multi-dimensional factor structures (see Figure 1). Specifically, the following five models were tested across the five data sets: Model 1) a single factor (general ODD); Model 2) a model with two orthogonal symptom factors (irritability versus oppositional behavior (OB)); Model 3) a model with 2 correlated symptom factors (irritability and OB) ; Model 4) a bifactor model with a general ODD factor and two orthogonal specific factors (irritability and OB); and finally Model 5) a modified bifactor model with general ODD and two correlated specific factors.

Within this hierarchical structured modeling approach, we also contrasted two primary competing models of irritability within ODD: one identifying irritability using the items touchy, angry and spiteful (e.g. Burke, et al., 2010) and the remaining ODD items loading on the oppositional behavior (OB) factor, and a second with the items temper, touchy and angry (e.g Stringaris & Goodman, 2009b) loading on irritability and the remainder on OB. Models were estimated using Mplus version 7 (Muthén & Muthén, 2010). In the twin study data sets, non-independence between twins within twin pairs was addressed using the clustering option. The weighted least squares means and variance (WLSMV) estimator was used given that this has

been shown to be optimal when data are ordinal (Bauer & Curran, 2004).

Model goodness of fit was evaluated using multiple indices, including the chi-square value, the Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and the weighted root mean square residual (WRMR; Loehlin, 2004). The acceptability of model fit was based on guidelines suggested in the literature: TLI $\geq$ 0.95 for excellent fit (Hu & Bentler, 1999) and TLI values between 0.90 - 0.95 for acceptable model fit (Bentler, 1990); RMSEA $\leq$ 0.08 for adequate fit and RMSEA $\leq$ 0.05 for close fit (Browne & Cudeck, 1993); WRMR $\leq$ 1.00 for good model fit (Yu, 2002). Overall adequacy of model fit was determined by considering all of the fit indices collectively.

Comparisons of alternative nested models were evaluated using the chi-square difference test. In some cases, non-linear constraints or theta parameterization were required for a model to converge. In these instances, it was not possible to conduct model comparisons using chi-square difference testing. These are noted in Table2.

Results

Across all data sets, the general single factor model of ODD symptoms (Model 1) showed indices of poor fit. Model 2 specified irritability and OB as orthogonal factors, whereas Model 3 estimated the correlation between the irritability and OB factors. Relative to all other models, the poorest fit resulted from specifying the irritability and OB factors as orthogonal. Marked improvements in fit were observed when the irritability and OB factors were correlated in Model 3.

General bifactor models. Model 4 specified a General ODD factor along with specific orthogonal irritability and OB factors, whereas Model 5 specified General ODD and two

correlated irritability and OB factors. Across data sets, fit indices for Model 4 were generally desirable, but did not uniformly exceed the criterion values indicative of good fit. Indices for Model 5 across data sets did exceed criterion values, indicating excellent fit.

Model contrasts. In order to provide additional evidence to determine the best fitting model, specific contrasts between models were evaluated. Chi square difference tests were conducted to contrast the general model with a) the correlated irritability and OB factor models (Model 3), and 2) the general factor and two correlated factors models (Model 5). In all cases, the general factor model was inferior to each of the other two models (see Table 2).

Also consistent across each data set, chi-square difference tests of Model 3 compared to Model 5 suggested that the latter resulted in superior fit. Thus, the inclusion of a general factor in addition to two correlated factors led to a significantly better fitting model. Additionally, chi-square difference tests indicated significantly better fit for Model 5 versus Model 4 across all datasets. In other words, along with a general factor, correlated specific factors showed significantly better model fit than orthogonal specific factors.

Comparing alternative models of ODD irritability. The two primary alternative models of irritability within ODD symptoms differ in the inclusion of one symptom. We will refer to the model which includes the symptoms of being touchy, angry and spiteful (Burke, 2012) as the Irr-Spite model, and the model which includes the items of touchy, angry and temper (Stringaris & Goodman, 2009b) as the Irr-Temper model.

In four of the five data sets, superior fit indices for Model 5 resulted from the Irr-Temper model. The exception was the PGS, in which the Irr-Spite model was superior.

Individual item parameters. Table 3 provides the item loadings from the best fitting model (Model 5) for the General factor and the two correlated specific factors from each data set.

On the General factor, for three of the five data sets (ALSPAC, PYS and PGS) the item loadings are uniformly significant and all are greater than .47. However, for the TTS and GTS, this is not the case. In the GTS, the items temper, argues and defies did not load significantly on the General factor, and in the TTS, none of the items loaded significantly on the general factor.

On the Specific Factors, all item loadings were significant for the GTS, TTS and ALSPAC, although in ALSPAC the magnitude of several loadings was modest; in ALSPAC the items temper, argues and defies each loaded below .43 on their respective factors. For the PGS, significant loadings were observed for all items except for annoys and blames, and the loadings for touchy and spiteful (on the irritability factor) and for temper (on the OB factor) were fairly low. Finally, in the PYS, only the temper item loaded significantly on the irritability factor, and the items of blames and spiteful did not load significantly on the OB factor.

Specific factor correlations. The irritability and OB factors for Model 5 were significantly correlated with one another in each data set. As shown in the final row of Table 3, these values range between .81 and .91 for all studies except PGS (.37).

Discussion

The adequate measurement of an irritability dimension within ODD has important implications for understanding comorbidity in models of developmental psychopathology, for the clinical identification of problems of irritability, and for the evaluation of new measures and diagnostic categories based on the manifestation of irritability in children. The goals of the present analyses were to systematically test the underlying structure of ODD symptoms, and to test competing models of irritability.

The results provide compelling evidence for multidimensionality within ODD and for the

superiority of a model of irritability based on the presence of the symptoms of often losing temper, being touchy, and being angry. This configuration of irritability was consistent across four large data sets. The data sets included a mixture of features: single and mixed gender samples, twin samples, varied ages of children, population representative samples and samples that were selected from the community to be disproportionately reflective of higher risk conditions. Further, each data set included a different parent-report measure of ODD, enhancing the impressive consistency in the replication of the factor structure across data sets. These results support the model of irritability represented in a recently introduced measure (Stringaris et al., 2012), which taps loss of temper, anger and being annoyed by others as key indicators.

The second major finding – replicated across all five data sets - was that ODD symptoms are best modeled using a bifactor model, including both a general ODD factor and two specific dimensions. This may mean that ODD is best understood as a construct with two symptom dimensions, but cannot meaningfully be severed into separate and independent components. Whilst recognizing that correlations between latent variables are typically higher than between observed variables, the factors correlations between the oppositional and headstrong factors was substantial (between 0.81 and 0.91) in all data sets, apart from the PGS, where the two factors correlated at .37. Modest to marked correlations between irritability and oppositional or headstrong behavioral factors have also been reported elsewhere, such as .89 (Aebi et al., 2010), .73 (Kreiger et al., 2013), or between .37 and .58 across alternative models (Ezpeleta et al., 2012). Furthermore, the model specifying two orthogonal factors in the absence of a general ODD factor yielded incontrovertibly the worst fit of any of the models. These results suggest that although two separate irritability and oppositional behavior factors capture unique variance, they should not be seen as independent from one another; a conclusion consistent with that

drawn by Ezpeleta and colleagues (2012).

Consistent with Martel and colleagues (2010), we suggest that the bifactor model supported here identifies related but distinct and meaningful constellations of symptoms within ODD that serve as etiological targets, that highlight differential treatment needs, and that indicate differential prognostic risks. For example, there is evidence of differential genetic etiology for the two dimensions, with irritability and depression overlapping due to shared genes in contrast to separate genetic explanation for the overlap between oppositional behavior and delinquency (Stringaris, Zavos, Leibenluft, Maughan, & Eley, 2012).

While our data demonstrate a bifactor configuration for ODD items, we note that some factor loadings are non-significant on either the general or specific factors, and that these differences vary across studies. For instance, in the TTS and GA samples, not all items loaded on the General factor. In the PYS, only "touchy" loaded on irritability, and only "spiteful" on OB. There are several possible explanations for these findings. First, it may be that the differences in loadings arise from the differences in measurement methods across the studies. Although all 5 studies inquired about 8 ODD items, the item response options and their scaling differed across studies. To test for equivalence of (or differences between) item loadings, it is necessary to have a similar scaling to the items. Because the scaling is different, this would be an important factor contributing to difference in loading magnitude across studies. The most restrictive example is the PYS data which had dichotomous response options in contrast to the more dimensional rating scales used in the other studies; this sample also had the most discrepant values for individual item loadings.

Other explanations are conceptual and relate to the available items. Eight ODD items may not be sufficient to adequately capture these constructs, particularly the irritability construct

which is measured using three items. Our *a priori* decision to restrict the number of dimensions in these models to two may have also led to some items having weaker loadings. For instance, the non-significant loading of the items annoys others and blames others in the PGS is consistent with exploratory factor analyses previously conducted with those data (Burke, et al., 2010), which suggested those two items in particular loaded on a separate factor from irritability and oppositional behavior altogether. In addition, given that some datasets in this study had only one item by which to measure a Hurtful dimension, we were not able to test models including this (e.g. Whelan, et al., 2013) as a third factor across these data sets. A final consideration is that there may be sex differences in the individual items as ODD indicators, given that temper, angry and annoyed did not identify irritability in the Pittsburgh Girls Study.

In short, these results demonstrate that 8 ODD items conform to a bifactor structure that is consistent across multiple samples with different participant characteristics as well as different ODD item response options. We cannot confirm that individuals respond to the ODD items in a similar way as explicit tests of metric invariance were not conducted. Such testing is needed prior to performing comparisons of individual factor loadings or item scores between samples.

In order to fully evaluate these issues and alternatives, further analyses using datasets with consistent item scaling, which also include multiple indicators of a hurtful (e.g. separate items for spiteful versus vindictive behaviors) dimension would be necessary. The need for this work is further highlighted by inconsistent evidence in support of models which include three symptom dimensions. For instance, although Stringaris and Goodman (2009a) found evidence that a Hurtful dimension of symptoms was differentially (in contrast to a Headstrong dimension) related to aggressive versus non-aggressive conduct problems, Krieger and colleagues (2013) found only modest evidence of divergent validity for separate Headstrong versus Hurtful

dimensions. Further, Whelan and colleagues (2013) found that the Headstrong dimension measured from late childhood through adolescence predicted both conduct problems and callousness at age 16, whereas the Hurtful dimension distinctly predicted none of the measured outcomes. Relatedly, although a separate dimension of Antagonistic behavior was supported in factor analyses of the Pittsburgh Girls Study (Burke et al., 2010), it showed no distinction from Oppositional Behavior in terms of predicting outcomes. Thus the literature suggests the possibility that a third dimension of symptoms exists, and that it may represent some heterogeneity of antisocial behavior, but is unclear whether such a dimension is functionally different from the first behavioral dimension.

Despite these caveats, and the need for further analyses, the present results were sufficient to address the primary questions of interest regarding irritability in the underlying structure of dimensions within ODD symptoms. They support the distinction of a narrow irritability dimension relative to a behavioral dimension, and the results suggest a preferred set of indicators for that irritability dimension. The results provide evidence that there is common variance shared by irritability and behavioral items – reflected in the general factor in the best-fitting models in all five data sets.

These results have implications for the new DSM 5 disorder category of Disruptive Mood Dysregulation Disorder (DMDD;American Psychiatric Association, 2013). DMDD posits that a significant group of youth may be identified by the presence of non-episodic irritability along with recurrent tantrum behaviors. The perception that many children with irritability were being misidentified as meeting criteria for bipolar disorder, and consequently being prescribed medications for that disorder, led to the development of DMDD (Copeland, Angold, Costello, & Egger, 2013; Stringaris, 2013). The defining features of DMDD appear to overlap markedly

with ODD-based irritability. Each includes chronic anger and persistent irritability as core features, and frequent temper outbursts are an indicator of irritability in the better fitting models in the present data. The primary differences between DMDD and the irritability dimension of ODD is that the criteria for DMDD specify a frequency of three or more temper outbursts per week, that DMDD is shown in at least two settings, and specify a duration of 12 months, rather than the 6 months duration required for ODD symptoms. DSM 5 indicates that when criteria for both disorders are met, a diagnosis of DMDD and not ODD should be given. This prohibition, and the broader implication that the irritability component can be, at some level of severity, divorced from oppositional or defiant behavior, does not yet have any empirical basis. In fact, analyses in multiple samples of youth have identified marked difficulty differentiating DMDD from behavioral disorders, especially ODD (Axelson, et al., 2012; Copeland, et al., 2013), calling into question the clinical utility of a diagnostic condition typified by irritability independent from oppositional behavior.

The present results would suggest further caution against a diagnostic categorization like DMDD, given the underlying structure of these symptoms identified in these analyses. This approach appears to be akin to the poorest fitting models in these analyses; those in which irritability was treated as independent from both oppositional behavior and the common features of general ODD. Nonetheless, it should be noted that the data in the present study did not include items assessing the frequency of tantrums, of symptoms in multiple contexts, or of a duration of 12 months. Thus, some caution is warranted when making inferences regarding DMDD based on the present results.

Limitations. As noted, although the variability of the questionnaires and response scaling across data sets was a strength of the study, it did also limit the ability to model some possible

alternatives. Furthermore, differences in item scaling leave some questions about the meaning of poor fit indicators for some items in some studies. In addition, the varying gender composition across studies raised some questions about potential variability in structure for girls versus boys. Despite these limitations, the study was able to assess the configural invariance of irritability symptoms, providing strong support for a model of two distinct but correlated dimensions, and for modeling irritability among ODD symptoms using the items of losing temper, being often touchy and being often angry.

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. (2013). *Diagnostic and Statistical Manual of Mental Disorders (Fifth ed.)*. Arlington, VA: American Psychiatric Publishing.

Axelson, D., Findling, R. L., Fristad, M. A., Kowatch, R. A., Youngstrom, E. A., Horwitz, S. M., Arnold, L. E., Frazier, T.W., Ryan, N., Demeter, C., Gill, M.K., Hauser-Harrington, J. C., Depew, J., Kennedy, S. M., Gron, B. A., Rowles, B. M. & Birmaher, B. (2012). Examining the Proposed Disruptive Mood Dysregulation Disorder Diagnosis in Children in the Longitudinal Assessment of Manic Symptoms Study. *Journal of Clinical Psychiatry, 73*(10), 1342-1350.

Bados, A., Gomez-Benito, J., & Balaguer, G. (2010). The State-Trait Anxiety Inventory, Trait Version: Does It Really Measure Anxiety? *Journal of Personality Assessment, 92*(6), 560-567.

Bezdjian, S., Krueger, R. F., Derringer, J., Malone, S., McGue, M., & Iacono, W. G. (2011). The structure of DSM-IV ADHD, ODD, and CD criteria in adolescent boys: A hierarchical approach. *Psychiatry Research, 188*(3), 411-421.

Boylan, K., Vaillancourt, T., Boyle, M., & Szatmari, P. (2007). Comorbidity of internalizing disorders in children with oppositional defiant disorder. *European Child & Adolescent*

Psychiatry, 16(8), 484-494.

Brotman, M. A., Rich, B. A., Guyer, A. E., Lunsford, J. R., Horsey, S. E., Reising, M. M., Thomas, L. A., Fromm, S. J., Towbin, K., Pine, D. S. & Leibenluft, E. (2010). Amygdala Activation During Emotion Processing of Neutral Faces in Children With Severe Mood Dysregulation Versus ADHD or Bipolar Disorder. *American Journal of Psychiatry, 167*(1), 61-69.

Brotman, M. A., Schmajuk, M., Rich, B. A., Dickstein, D. P., Guyer, A. E., Costello, E. J., Egger, H. L., Angold, A., Pine, D. S. & Leibenluft, E. (2006). Prevalence, clinical correlates, and longitudinal course of severe mood dysregulation in children. *Biol Psychiatry, 60*(9), 991-997.

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*(4), 319-326.

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

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*(5), 484-492.

Burke, J. D., Loeber, R., Lahey, B. B., & Rathouz, P. J. (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. *J Child Psychol Psychiatry, 46*(11), 1200-1210.

Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A

confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology, 29*(4), 339-349.

Carlson, G. A., & Cantwell, D. P. (1980). Unmasking masked depression in children and adolescents. *The American journal of psychiatry*.

Copeland, W. E., Angold, A., Costello, E. J., & Egger, H. (2013). Prevalence, Comorbidity, and Correlates of DSM-5 Proposed Disruptive Mood Dysregulation Disorder. *American Journal of Psychiatry, 170*(2), 173-179.

Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of general psychiatry, 66*(7), 764-772.

Costello, A. J., Edelbrock, C., Dulcan, M. K., Kalas, R., & Klaric, S. (1987). *Diagnostic Interview Schedule for Children (DISC)*. Pittsburgh, PA: Western Psychiatric Institute and Clinic.

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

Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Vanhorn, Y., Christ, M. A. G., Hart, E. A., Hanson, K. (1993). Oppositional Defiant Disorder and Conduct Disorder - a Meta-Analytic Review of Factor-Analyses and Cross-Validation in a Clinic Sample. *Clinical Psychology Review, 13*(4), 319-340.

Gadow, K. D., & Sprafkin, J. (1994). *Child Symptom Inventories manual*. Stony Brook, NY: Checkmate Plus.

Gibbons, R. D., Rush, A. J., & Immekus, J. C. (2009). On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. *Journal of Psychiatric Research, 43*(4), 401-410.

Hartman, C. A., Hox, J., Mellenbergh, G. J., Boyle, M. H., Offord, D. R., Racine, Y., McNamee, J., Gadow, K. D., Sprafkin, J., Kelly, K. L., Nolan, E. E., Tannock, R., Schachar, R., Schut, H., Postma, I., Drost, R. & Sergeant, J. A. (2001). DSM-IV internal construct validity: When a taxonomy meets data. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(6), 817-836.

Helzer, J. E., Kraemer, H. C., & Krueger, R. F. (2009). *Dimensional approaches in diagnostic classification: refining the research agenda for DSM-V*: American Psychiatric Pub.

Hipwell, A. E., Loeber, R., Stouthamer-Loeber, M., Keenan, K., White, H. R., & Kroneman, L. (2002). Characteristics of girls with early onset disruptive and antisocial behaviour. *Criminal behaviour and mental health, 12*(1), 99-118.

Hipwell, A. E., Stepp, S., Feng, X., Burke, J., Battista, D. R., Loeber, R., & Keenan, K. (2011). Impact of oppositional defiant disorder dimensions on the temporal ordering of conduct problems and depression across childhood and adolescence in girls. *Journal of Child Psychology and Psychiatry, 52*(10), 1099-1108.

Keenan, K., Hipwell, A., Chung, T., Stepp, S., Stouthamer-Loeber, M., Loeber, R., & McTigue, K. (2010). The Pittsburgh Girls Study: overview and initial findings. *Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 39*(4), 506-521.

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 Oppositionalay in a Brazilian Communiiy Sample: Testing the DSM-5 Proposal and Etiological Links. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(4), 389-400.

Kuny, A. V., Althoff, R. R., Copeland, W., Bartels, M., Van Beijsterveldt, C. E. M., Baer, J., & Hudziak, J. J. (2013). Separating the Domains of Oppositional Behavior: Comparing Latent Models of the Conners' Oppositional Subscale. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(2), 172-183.

Lavigne, J. V., Gouze, K. R., Bryant, F. B., & Hopkins, J. (2014). Dimensions of oppositional defiant disorder in young children: Heterotypic continuity with anxiety and depression. *Journal of abnormal child psychology*, 1-15.

Martel, M. M., Gremillion, M., Roberts, B., von Eye, A., & Nigg, J. T. (2010). The structure of childhood disruptive behaviors. *Psychological assessment, 22*(4), 816-826.

Martel, M. M., von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: is there a 'g' factor? *Journal of Child Psychology and Psychiatry, 51*(8), 905-914.

Moffitt, T. E., Arseneault, L., Jaffee, S. R., Kim-Cohen, J., Koenen, K. C., Odgers, C. L., Slutske, W. S., & Viding, E. (2008). Research Review: DSM-V conduct disorder: research needs for an evidence base. *Journal of Child Psychology and Psychiatry, 49*(1), 3-33.

Muthén, L. K., & Muthén, B. (2010). *Mplus*: Muthén & Muthén Los Angeles, CA.

Patrick, C. J., Hicks, B. M., Nichol, P. E., & Krueger, R. F. (2007). A bifactor approach to modeling the structure of the psychopathy checklist-revised. *Journal of Personality Disorders, 21*(2), 118-141.

Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM-

III-R symptoms for the disruptive behavior disorders. *J Am Acad Child Adolesc Psychiatry, 31*(2), 210-218.

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research, 47*(5), 667-696.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19-31.

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), 726-738.

Rowe, R., Maughan, B., & Eley, T. C. (2006). Links between antisocial behavior and depressed mood: The role of life events and attributional style. *Journal of Abnormal Child Psychology, 34*(3), 293-302.

Simms, L. J., Gros, D. F., Watson, D., & O'Hara, M. W. (2008). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and anxiety, 25*(7), E34-46.

Stringaris, A. (2013). Editorial: The new DSM is coming - it needs tough love. *Journal of child psychology and psychiatry, and allied disciplines, 54*(5), 501-502.

Stringaris, A., Cohen, P., Pine, D. S., & Leibenluft, E. (2009). Adult outcomes of youth irritability: a 20-year prospective community-based study. *Am J Psychiatry, 166*(9), 1048-1054.

Stringaris, A., & Goodman, R. (2009a). Longitudinal outcome of youth oppositionality: irritable, headstrong, and hurtful behaviors have distinctive predictions. *J Am Acad Child Adolesc Psychiatry, 48*(4), 404-412.

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

Stringaris, A., Goodman, R., Ferdinando, S., Razdan, V., Muhrer, E., Leibenluft, E., & Brotman, M. A. (2012). The Affective Reactivity Index: a concise irritability scale for clinical and research settings. *Journal of Child Psychology and Psychiatry, 53*(11), 1109-1117.

Stringaris, A., Zavos, H., Leibenluft, E., Maughan, B., & Eley, T. C. (2012). Adolescent Irritability: Phenotypic Associations and Genetic Links With Depressed Mood. *American Journal of Psychiatry, 169*(1), 47-54.

Whelan, Y. M., Stringaris, A., Maughan, B., & Barker, E. D. (2013). Developmental Continuity of Oppositional Defiant Disorder Subdimensions at Ages 8, 10, and 13 Years and Their Distinct Psychiatric Outcomes at Age 16 Years. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(9), 961-969.

Table 1

Sample Demographics and ODD Symptom Measure Used

Sample	**N**	**Gender**	**Ages**	**Measure**
Pittsburgh Youth Study	1,517	Boys	7, 10, 13	DISC
Pittsburgh Girls Study	2,451	Girls	5-8	CSI-4
The Avon Longitudinal Study of Parents and Children (ALSPAC)	7,420	Boys and Girls	10-11	DAWBA
Tennessee Twins Study	4,046	Boys and Girls	6-17	CAPS
Georgia Twin Study	1,692	Boys and Girls	6-18	ERDS

Notes. ODD = oppositional defiant disorder. DISC = Diagnostic Interview Schedule for Children. CSI-4 = Child Symptom Inventory – 4. DAWBA = Developmental and Well-Being Assessment. CAPS = Child and Adolescent Psychopathology Scale. ERDS = Emory Diagnostic Rating Scale.

Table 2. CFAs of competing irritability models across five large data sets.

	Irr-Spite						Irr-Temper					
Dataset: ALSPAC												
Model	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p
1	344.2	.989	.067*	2.07	NA		344.24	.989	.067	2.07	246.4	<.0001
2	[T]24,488.0	.156	.578*	28.50	NA		22,773	.215	.557*	28.3	15953.0	<.0001
3	247.4	.992	.057*	1.72	NA		233.0	.992	.055	1.61	160.6	<.0001
4	91.6	.995	.043	0.94	NA		100.77	.995	.045	0.95	45.7	<.0001
5	[C]86.0	.995	.043	0.86			**37.5**	**.998**	**.026**	**0.54**		
Dataset: TTS												
Model	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p
1	555.39	0.97	.088*	2.49	343.40	<.0001	555.39	.974	.088*	2.49	438.38	<.0001
2	8,646.67	0.21	.486*	17.89	5856.35	<.0001	6934.22	.290	.460*	17.05	5333.61	<.0001
3	537.88	.973	.090*	2.46	321.84	<.0001	310.14	.986	.065*	1.79	211.22	<.0001
4	[T]249.80	.971	.066	1.33	NA		95.09	.990	.043	0.92	34.47	<.0001
5	202.71	.98	.069	1.33			**64.01**	**.996**	**.036**	**.75**		
Dataset: PYS												
Model	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p
1	68.14	.987	.04	1.09	NA		68.14	.987	.04	1.09	51.17	<.0001
2	1,569.54	.367	.294	8.07	NA		1,980.47	.200	.33	9.05	1628.15	<.0001
3	56.02	.990	.037	0.98	NA		62.16	.988	.04	1.04	45.57	<.0001
4	[T]26.43	.993	.026	.641	NA		37.31	.990	.037	.771	19.90	<.0001
5	[T]20.98	.995	.022	.564			**15.19**	**.998**	**.016**	**.474**		
Dataset: GTS												
Model	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p
1	395.1	.97	.103*	1.72	NA		395.1	.97	.103*	1.72	265.31	<.0001
2	11,692	.06	.575*	18.19	NA		12,594	-.01	.597*	18.74	8,752.71	<.0001

3	327.7	.974	.096*	1.55	NA		349.1	.972	.099*	1.60	221.70	<.0001
4	^{T}157.4	.980	.083*	10.99	NA		^{T}176.8	.978	.088*	1.06	NA	
5	^{T}168.0	.977	.090	0.99			**102.5**	**.987**	**.069***	**0.74**		
Dataset: PGS												
Model	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p	χ^2	TLI	RMSEA	RMR	$\Delta\chi^2$	p
1	527.6	.921	.103*	2.51	68.99	<.0001	527.6	.921	.103*	2.51	311.06	<.0001
2	5,047.8	.217	.320	10.70	1547.08	<.0001	5,901.5	.084	.346*	11.6	2,365.63	<.0001
3	432.7	.932	.094	2.26	45.53	<.0001	509.6	.920	.103*	2.46	292.48	<.0001
4	111.1	.974	.058	1.06	6.26	.012	205.7	.950	.081*	1.48	9.91	.002
5	**71.2**	**.983**	**.047**	**.88**			113.8	.971	.068	1.05		

Notes: Bold font = best fitting model within data set. * p-value for hypothesis that RMSEA ≤ .05; T = Theta parameterization used due to negative residual variance; C = non-linear constraints required to achieve model fit. NA = not available: in cases where non-linear constraints or theta parameterization were required to fit one of the models to be compared, these models were excluded from chi-square difference testing.

Models: 1 = 1 General Factor; 2 = 2 orthogonal ODD factors (IRR + OB); 3 = 2 correlated ODD factors (IRR + OB); 4 = 1 General Factor + 2 orthogonal ODD factors; 5 = 1 General Factor + 2 correlated ODD factors; Irr-Spite = touchy, angry and spiteful, Irr-Temper = touchy, angry, loses temper.

Table 3. ODD item loadings from Model 5 for each data set.

General Factor Loadings															
	ALSPAC			TTS			GTS			PYS			PGS*		
ODD Item	β	se	p	β	se	p	β	se	p	B	se	p	β	se	p
1. temper	.81	.06	<.001	.17	.26	.52	-.34	.27	.21	.60	.16	.00	.61	.03	<.001
2. argues	.91	.06	<.001	.57	.29	.05	-.10	.38	.79	.72	.06	.00	.49	.04	<.001
3. defies	.83	.04	<.001	.12	.31	.71	-.29	.31	.35	.80	.11	.00	.50	.03	<.001
4. annoys	.66	.11	<.001	-.14	.30	.65	-.54	.24	.02	.47	.19	.02	.72	.02	<.001
5. blames	.59	.12	<.001	-.06	.28	.85	-.52	.24	.03	.53	.18	.00	.68	.02	<.001
6. touchy	.71	.06	<.001	.05	.26	.85	-.57	.23	.01	.61	.13	.00	.60	.02	<.001
7. angry	.71	.10	<.001	-.01	.26	.96	-.58	.27	.03	.72	.10	.00	.60	.00	<.001
8. spiteful	.63	.11	<.001	-.24	.31	.43	-.66	.23	.003	.70	.05	.00	.69	.02	<.001

Specific Irritability and Oppositional Behavioral Factor Loadings																
	ALSPAC			TTS			GTS			PYS				PGS*		
ODD Item	β	se	p	β	se	p	β	se	p	β	se	P	ODD Item	β	se	p
	Irritability Dimension															
1. temper	.39	.12	<.001	.79	.06	<.001	.75	.11	<.001	-.60	.11	.00	6. touchy	.24	.05	<.001
6. touchy	.66	.10	<.001	.78	.02	<.001	.62	.21	<.001	-.16	.23	.48	7. angry	.75	.11	<.001
7. angry	.67	.11	<.001	.77	.01	<.001	.74	.22	<.001	-.34	.26	.19	8. spiteful	.26	.05	<.001
	Oppositional Behavioral Dimension															
2. argues	.31	.15	.04	.76	.22	.001	.96	.05	<.001	-.61	.15	.00	1. temper	.37	.04	<.001
3. defies	.42	.14	<.001	.80	.05	<.001	.76	.12	<.001	-.58	.17	.00	2. argues	.65	.04	<.001
4. annoys	.63	.11	<.001	.78	.06	<.001	.59	.21	.01	-.42	.19	.03	3. defies	.53	.04	<.001
5. blames	.71	.10	<.001	.73	.03	<.001	.62	.21	<.001	-.32	.23	.17	4. annoys	.03	.05	.52
8. spiteful	.63	.11	<.001	.82	.10	<.001	.58	.26	.03	-.15	.22	.51	5. blames	-.07	.05	.19
Correlation between Irritability and Oppositional Behavioral Dimension																
Correlation	.84	.06	<.001	.85	.02	<.001	.91	.05	<.001	.81	.21	<.001		.37	.09	<.001

Note: ODD = Oppositional defiant disorder. * PGS loadings result from the Irr-Spite model, in which the symptoms of touchy, angry and spiteful load on the Irritability dimension, while the remainder are the result of the Irr-Temper model (touchy, angry, loses temper).

Figure 1: Unidimensional and multidimensional factor models of irritability and oppositional behavior among oppositional defiant disorder