Published in final edited form as:

Assessment. 2018 September ; 25(6): 744–758. doi:10.1177/1073191116660381.

Externalizing Behavior Across Childhood as Reported by Parents and Teachers: A Partial Measurement Invariance Model

Kevin M. King¹, Jeremy W. Luk¹, Katie Witkiewitz², Sarah Racz³, Robert J. McMahon^{4,5}, Johnny Wu⁶, and the Conduct Problems Prevention Research Group^{*}

¹University of Washington, Seattle, WA, USA

²University of New Mexico, Albuquerque, NM, USA

³University of Maryland, College Park, MD, USA

⁴Simon Fraser University, Vancouver, British Columbia, Canada

⁵Child & Family Research Institute, Vancouver, British Columbia, Canada

⁶Johnson & Johnson, Jacksonville, FL, USA

Abstract

The externalizing spectrum may explain covariation among externalizing disorders observed in childhood and adulthood. Few prospective studies have examined whether externalizing spectrum might manifest differently across time, reporters, and gender during childhood. We used a multitrait, multimethod model with parent and teacher report of attention-deficit/hyperactivity disorder (ADHD) symptoms, oppositional defiant disorder (ODD) symptoms, and conduct disorder (CD) symptoms from kindergarten to Grade 5 in data from the Fast Track Project, a large multisite trial for children at risk for conduct problems ($n = 754$). The externalizing spectrum was stably related to ADHD, ODD, and CD symptoms from kindergarten to Grade 5, with similar contributions from parents and teachers. Configural, metric, and scalar invariance were largely supported across time, suggesting that the structure of the externalizing spectrum is stable over time. Configural and partial metric invariance were supported across gender, but scalar invariance was not supported, with intercepts consistently higher for males than for females. Overall, our findings confirm other research that the externalizing spectrum can be observed early in development as covariation between ADHD, ODD, and CD, and extend that work to show that it is relatively consistent across time and reporter, but not consistent across gender.

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Kevin M. King, Department of Psychology, University of Washington, Box 351525, Seattle, WA, 98195, USA. kingkm@uw.edu.

^{*}Members of the Conduct Problems Prevention Research Group include Karen L. Bierman (Pennsylvania State University), Kenneth A. Dodge (Duke University), Mark T. Greenberg (Pennsylvania State University), John E. Lochman (University of Alabama), Robert J. McMahon (Simon Fraser University and Child & Family Research Institute), and Ellen E. Pinderhughes (Tufts University).

Declaration of Conflicting Interests

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

Keywords

externalizing spectrum; developmental psychopathology; oppositional defiant disorder; conduct disorder; attention-deficit/hyperactivity disorder

Childhood externalizing behaviors co-occur substantially during childhood (e.g., Achenbach, 1978). Attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD), as well as symptoms of these disorders, have been shown to co-occur at rates far higher than chance (Essau, 2003; Waschbusch, 2002). In a meta-analytic review of general population studies, children with a *DSM* diagnosis of ADHD had 10 times the odds of being diagnosed with either CD or ODD relative to children without ADHD (Angold, Costello, & Erkanli, 1999), and a more recent review of the literature suggested that a comorbid diagnosis of CD or ODD was found in approximately half of ADHD cases (Connor, Steeber, & McBurnett, 2010).

The covariation among ADHD, ODD, and CD can be modeled as a single latent externalizing factor (Beesdo-Baum et al., 2009; Reitz, Dekovi, & Meijer, 2005; Wittchen et al., 2009), often referred to as the “externalizing spectrum” (Krueger et al., 2002). Behavioral genetic studies spanning childhood and adolescence suggested that there are substantial genetic and nonshared environmental influences on this externalizing factor (ages 9–11 years: Bornovalova, Hicks, Iacono, & McGue, 2010; Tuvblad, Zheng, Raine, & Baker, 2009; age 14 years: Dick, Viken, Kaprio, Pulkkinen, & Rose, 2005; Silberg et al., 1996; Thapar, Harrington, & McGuffin, 2001; ages 13–19 years: Knopik, Heath, Bucholz, Madden, & Waldron, 2009; Nadder, Rutter, Silberg, Maes, & Eaves, 2002, but cf. Burt, 2009). However, some authors have argued that spectrum models have not sufficiently considered *both* shared and unique variability in ADHD, CD, and ODD (Beauchaine & McNulty, 2013). For example, ADHD may progress to CD in the presence of coercive or ineffective parenting (Beauchaine, Hinshaw, & Pang, 2010), whereas negative emotionality may play a significant role in the development of ODD (Loeber, Burke, & Pardini, 2009). Taken together, the evidence suggests that a properly specified model of the externalizing spectrum will explicitly model both co-occurring and unique variability in childhood externalizing behaviors, which should better inform etiological models of their (co)development.

A Multitrait Model of Externalizing Behaviors

Multitrait models of ADHD, ODD, and CD symptoms have been explicitly modeled in a few recent studies (Bezdjian et al., 2011; Burns, de Moura, Beauchaine, & McBurnett, 2014; Olson et al., 2013). Using principal components analysis, Bezdjian et al. (2011) showed that, among adolescent boys, a single externalizing component explained about 24% of total variance, whereas a three-component (ADHD, ODD, CD) conceptualization increased the variance accounted for to 36%, suggesting the presence of variance in symptoms specific to each disorder. Burns et al. (2014) used a bifactor latent variable model to show that ADHD and ODD symptoms reflected both common (accounting for 70% across reporters), and ADHD-specific (4% to 14%) and ODD-specific (11%) variance. Other studies have shown similar results using bifactor models of personality pathology (such as general and relational

aggression and rule breaking) associated with externalizing behaviors (e.g., Tackett, Daoud, De Bolle, & Burt, 2013; Tackett, Herzhoff, Reardon, De Clercq, & Sharp, 2014). Lorber, Del Vecchio, & Slep (2014) showed that averaged maternal and paternal report of infant (age 8–24 months) externalizing behaviors (aggression, defiance, activity, and distress) were explained by both a single externalizing factor and by correlations among specific residual covariances across time.

Each of these prior studies has shown pieces of the whole, by illustrating either different types of multitrait externalizing models or combining multiple reporters of a child's externalizing behaviors, but each has limitations. For example, correlating uniquenesses (i.e., residual variances) may appropriately model disorder-specific covariation, but those uniquenesses cannot be used in predictive models (Smith, 2005). Bifactor models overcome this hurdle, but they have not been tested across development. Moreover, nearly all of this work has relied on parent report, but no prior studies have explicitly modeled reporter factors alongside common and disorder-specific externalizing factors. It remains unclear to what extent common and specific multitrait variance may be confounded with reporter variance. Parents and teachers, for example, may have differing perceptions of child's externalizing behaviors due to differences in the reference group to which they compare the child, and differences in the context in which they observe the child. For the current study, we tested a multitrait, multi-reporter factor model that captured variability in the common externalizing variance across ADHD, ODD, and CD symptoms (i.e., the externalizing spectrum); variability unique to ADHD, ODD, and CD symptoms; and variability attributable to different reporters (e.g., Eid et al., 2008; Eid, Lischetzke, Nussbeck, & Tielweiler, 2003).

Invariance of the Externalizing Spectrum Across Development

There is substantial evidence that externalizing behaviors can have a different behavioral expression across development (Beauchaine et al., 2010; Patterson, 1993). For example, what may be a somewhat normative behavior in kindergarten, such as talking out of turn or having trouble sitting still, would be expected to be a very nonnormative behavior by sixth grade. If the externalizing spectrum reflects a general liability to externalizing behaviors (Krueger et al., 2002), we might expect a consistent association between the externalizing spectrum and its indicators over time, such as ADHD, CD, and ODD. Multiple studies have posited a developmental sequence (i.e., “developmental cascade”) that begins with the early onset of ADHD, evolving into early childhood ODD and later childhood CD, which then progresses into continued delinquency, substance use, and antisociality from adolescence into adulthood (e.g., Beauchaine et al., 2010; Dodge et al., 2009; Loeber & Burke, 2011; Loeber, Burke, Lahey, Winters, & Zera, 2000; Martel et al., 2009; Molina, 2011). Furthermore, some research showed that most children with comorbid ADHD and ODD did not progress to CD at follow-up, particularly if genetic and environmental influences on CD are controlled (e.g., Biederman et al., 2008; Diamantopoulou, Verhulst, & van der Ende, 2011; Lahey et al., 2009). If some behavior problems (such as CD or ODD) are more typically observed later in development, it may be that their relation to the latent externalizing spectrum changes as well, increasing over time as these disorders become more prominent.

Measurement invariance (Meredith, 1993; Vandenberg & Lance, 2000) may be used to test whether the components of a latent factor (such as the indicators, their loadings and intercepts, or factor means and variances) are equal across groups or time. Without measurement invariance, trait scores are not comparable across ages, and it would be impossible to disaggregate true changes in externalizing behavior scores over time from changes produced by changes in the underlying structure of those scores, or even simply compare externalizing behavior symptom scores over time. Two prior studies examined factor loading invariance (i.e., equivalent factor loadings over time) of the externalizing spectrum model across development, but neither of these studies tested intercept invariance (i.e., equivalent means in disorder-specific symptoms) of the externalizing spectrum across development. First, Young et al. (2009) identified an externalizing spectrum model with ADHD, CD, substance use, and novelty seeking among adolescents, and found that factor loadings for ADHD were stable across time points, whereas factor loadings for both CD and substance use increased from ages 12 to 17 years. However, the authors combined parent and teacher report of symptoms across ages. Although this may provide a more stable estimate of symptoms, without establishing invariance across time may compound measurement error and provide inaccurate estimates of means as well as biasing regression estimates (Millsap, 1997). Moreover, failing to separately model reporter effects is likely to produce inaccurate estimates of the loadings of the indicators with the latent factor, as variance attributable to the trait is confounded with reporter-specific variance. Recently, Lorber et al. (2014) showed noninvariant factor loadings in a latent externalizing factor from 8 to 24 months, suggesting that there is early instability in the structure of the externalizing spectrum. However, no study that we are aware of has tested both loading and intercept invariance across multiple time points during middle childhood.

Gender Differences in the Externalizing Spectrum Model

Finally, there are well-characterized differences across gender in the expression of externalizing symptoms. Epidemiological studies have shown that boys are more likely than girls to be diagnosed with externalizing disorders (see Merikangas, Nakamura, & Kessler, 2009, for a review). For example, in the Great Smokey Mountain Study, boys were diagnosed with higher rates of ADHD (1.5% for boys, 0.3% in girls), ODD (3.1% for boys, 2.1% in girls), and CD (4.2% for boys, 1.2% in girls; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). Differences in the prevalence of behavior disorders across groups do not necessarily imply that the associations among these disorders would also differ. To date, only a few studies have empirically tested measurement invariance across gender in childhood externalizing behaviors, and none have tested gender differences over time. Burns, Walsh, Gomez, and Hafetz (2006) reported configural, metric, and scalar invariance on ADHD and ODD symptoms across gender in a large sample of American children and a parallel sample of Malaysian children. Moreover, they showed that American boys had a higher factor mean in both ADHD and ODD symptoms relative to American girls. On the other hand, using data from the current sample, we previously found scalar noninvariance in a brief kindergarten measure of behavior problems (a 10-item version of the Teacher Observation of Classroom Adaptation–Revised, TOCA-R) across gender, indicating that for the same overall level of behavior problems boys were rated as exhibiting more overt

behavior problems, whereas girls were rated as having more covert behavior problems (Wu, King, Witkiewitz, McMahon, & the Conduct Problems Prevention Research Group, 2012). Given conflicting reports as to whether measures that tap constructs along the externalizing spectrum are invariant across gender or not, it is of interest to test whether the structure of the externalizing spectrum itself is invariant across gender.

In light of increasing emphasis on understanding etiological factors that precede and predict the onset of externalizing behaviors, it is important to examine the structure of these behaviors and how that structure may change across development. Identifying the degree to which childhood externalizing behaviors do and do not share variance across multiple ages in childhood facilitates the discovery of more precise predictors of both externalizing behaviors and disorder-specific factors. Finally, there are few and conflicting reports as to whether measures that tap externalizing behaviors are or are not invariant across age or gender. Thus, the three goals of the current study were to (a) test a multitrait, multimethod model of externalizing behaviors across middle childhood, (b) to test for measurement invariance across time, and (c) gender.

Method

Participants and Design

The current study used data from a community-based sample of children drawn from the Fast Track project, a longitudinal multisite investigation of the development and prevention of childhood conduct problems (Conduct Problems Prevention Research Group, 1992, 2000). Schools within four sites (Durham, NC; Nashville, TN; rural Pennsylvania; and Seattle, WA) were identified as high risk based on crime and poverty statistics of the neighborhoods that they served. Within each site, schools were divided into sets matched for demographics (size, percentage free or reduced lunch, ethnic composition), and the sets were randomly assigned to control and intervention groups. Using a multiple-gating screening procedure that combined teacher and parent ratings of disruptive behavior, 9,594 kindergarteners across three cohorts (1991–1993) from 55 schools were screened initially for classroom conduct problems by teachers, using the Teacher Observation of Child Adjustment–Revised Authority Acceptance scale (Werthamer-Larsson, Kellam, & Wheeler, 1991). Those children scoring in the top 40% within cohort and site were then solicited for the next stage of screening for home behavior problems by the parents, using items from the Child Behavior Checklist (CBCL; Achenbach, 1991a) and similar scales, and 91% agreed ($n = 3,274$). The teacher and parent screening scores were then standardized and summed to yield a total severity-of-risk screen score. Children were selected for inclusion into the study based on this screen score, moving from the highest score downward until desired sample sizes were reached within sites, cohorts, and groups. Deviations were made when a child failed to matriculate in the first grade at a core school ($n = 59$) or refused to participate ($n = 75$), or to accommodate a rule that no child would be the only girl in an intervention group. The outcome was that 891 children (446 control and 445 intervention) participated. The current study excluded children from the intervention group. In addition to the high-risk sample, a stratified normative sample of 387 children (79 overlapped with the high-risk control group) was identified to represent the population normative range of risk scores and

was followed over time. The total number of individuals considered in the current study was 308 normative and 446 high-risk (control; $N = 754$). Participants were approximately 5 years old during the initial assessment at kindergarten and follow-up assessments were conducted annually through 2 years post-high school (approximately age 20 years). Across time, an average of 90% of participants were retained at each time point, and prior analyses of these data suggested that participants lost to follow-up did not significantly differ from those retained (Conduct Problems Prevention Research Group, 1999). Although we used a school-based sampling strategy, and thus participants were initially clustered by school in Grade 1, over time there was substantial change in this degree of clustering as children moved and transferred schools (Conduct Problems Prevention Research Group, 2002). Thus, accounting for clustering was not necessary in the current analyses. The sample was 47% White and 51% African American (with the remaining 2% representing other ethnicities), and was 69% male (because of oversampling for conduct problems). We used data from childhood assessments at Grades K, 2, 4, and 5 that shared common measures of externalizing behaviors across childhood, described in more detail below. All procedures were approved by institutional review boards at each respective site.

Measures

Childhood Externalizing Spectrum—*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, Text Revision (*DSM-IV-TR*; American Psychiatric Association, 2000) reports of externalizing symptoms (CD, ODD, and ADHD) at all time points were obtained using parent (with the CBCL, Achenbach, 1991a) and teacher report (with the Teacher's Report Form, TRF; Achenbach, 1991b), using items from the *DSM*-oriented subscales described by Achenbach, Dumenci, and Rescorla (2003). This study showed that *DSM-IV-TR* symptoms of ADHD, ODD, and CD symptoms may be identified in both the CBCL and the TRF, as rated by expert child psychologists and psychiatrists. Further research has supported the validity of using CBCL items to screen for externalizing disorders (Hudziak, Copeland, Stanger, & Wadsworth, 2004; Krol, De Bruyn, Coolen, & van Aarle, 2006). The same versions of the CBCL and TRF were used at all time points. For CD, we used 13 items from the CBCL, such as bullying, fighting, and stealing, and 8 items from the TRF such as being cruel to others, destroying property, and being truant. For ODD, we used five items from the CBCL, such as being disobedient, arguing, and having temper tantrums, and eight items from the TRF, such as being impatient, stubborn, and disturbing others. For ADHD, we used 8 items from the CBCL such as not concentrating, avoiding schoolwork, daydreaming, not sitting still, talking too much, and impulsive, and 15 items from the TRF such as hyperactive, fidgets, and cannot concentrate. It is important to note that, to reduce participant burden, not all CBCL or TRF items were administered at all waves, thus we could not perfectly re-create the *DSM*-oriented scales of Achenbach et al. (2003). Second, that not all *DSM-IV-TR* symptoms were included for each disorder, which means that these scales may not capture the full range or diversity of presentation of these disorders, and that without the inclusion of other criteria (such as impairment) or diagnostic cutoffs, our measures of ADHD, ODD, and CD reflect measures of *symptoms*, but not of the presence/absence or degree of *disorder* (Achenbach et al., 2003).

At each time point, a within-reporter mean (i.e., of CBCL or TRF *DSM*-oriented items) of CD, ODD, and ADHD symptoms was computed. Thus, a total of six scales (three parent-report scales and three teacher-report scales) were used as indicators for subsequent factor models.

Statistical Analysis

We modeled the correlated trait-correlated method externalizing spectrum model using the Mplus statistical package (Version 6.0; Muthén & Muthén, 2010). To account for missing data, Mplus uses full-information maximum likelihood (FIML) with the expectation-maximization algorithm (EM, Allison, 2002) to obtain estimates with robust standard errors (the MLR estimator in Mplus). FIML and EM are accepted approaches to handling missing data (Little & Rubin, 2002) when data are missing at random. To account for the oversampling of high-risk children in the Fast Track project and to increase generalizability to the general population, we used a probability weight based on group (normative vs. high-risk control) that had been previously calculated for all normative and high-risk control group participants (see Jones, Dodge, Foster, Nix, & the Conduct Problems Prevention Research Group, 2002, for description of the creation and calculation of this weighting variable). In short, the probability weight was calculated such that the weighted sample approximated a normal distribution of the sampled population across screening variables. These weights were calculated based on the distributions of the TOCA-R (Werthamer-Larsson et al., 1991) and problem behavior scores on the CBCL (Achenbach, 1991a).

Model fit was assessed using chi-square as an indicator of exact fit. Where exact fit was not achieved (as chi-square is sensitive to violations of normality and sample size; Hu & Bentler, 1999), we used relative fit indices, including the Bayesian information criteria (BIC), comparative fit index (CFI) and root mean square error of approximation (RMSEA). Using these indices, we judged model fit with reference to standards provided by Hu and Bentler (1999), Kenny and McCoach (2003) and the cautions of Marsh, Hau, and Wen (2004), and attended to residuals and modification indices as a means of balancing different indicators of model fit (Jackson, Gillaspay, & Purc-Stephenson, 2009).

Analyses proceeded as follows. Because we were modeling seven trait factors (a latent externalizing factor at four time points, as well as trait ADHD, ODD, and CD factors), reported by two different reporters (parent and teacher), we used a variation of the correlated-trait, multimethod minus one (CT-CM-1) approach outlined by Eid and colleagues (Eid et al., 2003; Eid et al., 2008). Figure 1 illustrates the CT-CM-1 model. In this method, we estimated both general and symptom-specific externalizing traits and a single reporter factor (i.e., “minus one”). For model identification, each type of latent factor (general and specific externalizing factors and reporter factor) were forced to be *uncorrelated*. The four general externalizing trait factors were estimated at each time point (Grades K, 2, 4, and 5), and allowed to freely correlate with one another. This factor was indicated by parent and teacher report of ADHD, ODD, and CD symptoms, and modeled the common variance among externalizing symptoms, between reporters, within time. The symptom-specific externalizing factors were indicated by symptoms of ADHD, ODD, or CD across all time points, which modeled the common variance between reporters that was

specific to ADHD, ODD, or CD across time. Finally, we estimated a single reporter factor at each time point, indicated by either teacher or parent indicators of externalizing behaviors (teacher report of ADHD, ODD, and CD) within each time point. This latent factor models the variance within reporter, between symptoms and within each time point. For identification and to avoid overfactoring, only one reporter factor was included; this effectively means that the parent report is treated as the “baseline method.” In these models, method variance because of the baseline method is not explicitly modeled, and the latent externalizing factor and the specific ADHD, CD, and ODD factors represent the factors as measured by the baseline reporter, while the variance specific to the other reporter is captured by the loadings on the general and specific externalizing factors. Because of this, we tested two models, one with a parent-reporter factor, which treats teacher report as the baseline method, and a second with a teacher-reporter factor, treating parent report as the baseline method, and compared results to understand how each reporter influenced the findings. To the degree that models differed across the choice of baseline method, this implies that the baseline method influences the latent factor structure.

Measurement invariance testing across time involved four steps: (a) configural invariance, (b) metric invariance, (c) scalar invariance, and finally (d) mean and variance invariance. Configural invariance tests the degree to which the latent construct (the externalizing factor) is represented by the same indicators across time. Metric invariance tests whether the loadings are equal across time. Equivalence of loadings shows that the latent factors are related to each scale in the same way across time or group (Reise, Widaman, & Pugh, 1993). We tested for metric invariance for all latent factors (general and specific externalizing factors, as well as reporter factors). Scalar invariance requires equality constraints to be specified for all indicator intercepts across time or group. Intercepts are the values of the observed individual scales at the mean level of the underlying latent factor (Meredith, 1993). Finally, mean and variance invariance tests whether the latent factor means (i.e., the average level of the latent factors) and variances (i.e., the distribution of the latent factors) are equal over time. We used an omnibus method to test invariance, which constrained all parameters to be equal over time or group, and used model modification indices to determine which parameters may not be invariant, comparing model alternatives using BIC.

Results

Descriptive Statistics of Study Variables

Table 1 shows the means and reliabilities of each externalizing spectrum indicator during childhood. TRF ADHD, ODD, and CD scales showed good reliability over time (a range = .84 to .92), as did CBCL ODD and CD scales (a range = .75 to .82). Parent-reported ADHD symptoms had somewhat lower but acceptable reliability (a range = .63 to .73). In general, more externalizing symptoms were endorsed by teachers than parents at each time point. Both teachers and parents endorsed that more ODD than CD items were endorsed at each time point, but parent and teacher ratings of ODD and CD both tended to decrease over time.

The Childhood Externalizing Spectrum Over Time

We tested our hypotheses as follows: (a) Did a CT-CM-1 model explain variability in externalizing symptoms over time? (b) Was there measurement invariance in the externalizing spectrum across development?

We first tested a CT-CM-1 model of the externalizing spectrum based on parent and teacher report of ADHD, ODD, and CD across Grades K, 2, 4, and 5, first with parent report as the “baseline method” and a teacher-report factor (which we refer to as the “parent baseline model”), and then vice versa (referred to as the “teacher baseline model”). All loadings and indicators were freely estimated, and all latent factor variances were fixed to 1. Table 2 presents model fit statistics for all models. Table 3 presents the loadings, intercepts, and estimated variances for the final models, while latent factor correlations from these models are presented in Table 4. Model fit for parent and teacher baseline models for these initial model(s) was within the acceptable range. Teacher-reported CD and ODD symptoms had nonsignificant and near-zero loadings ($-.03$ to $.09$) onto their respective specific factors (for both the parent and teacher baseline models), so we fixed their loadings on the CD and ODD specific factors to zero. These adjusted initial models did not fit worse than the initial model. At each time point, this CT-CM-1 model explained a significant amount of variability in the indicators, with 66% to 84% of the variance in parent-report items explained across models ($M = 76\%$), and 30% to 78% of the variance in teacher report items explained across models ($M = 56\%$). This first model thus established configural invariance of the general externalizing spectrum over time.

This model suggests that variability in ADHD symptoms was driven by four sources of variance: general variance in externalizing behaviors, variability specific to ADHD, variability in general externalizing that is specific to the reporter (parent or teacher), and error. Conversely, variance in CD and ODD symptoms was driven by general externalizing spectrum variance, variance in CD or ODD symptoms that was specific only to parent report, variability in general externalizing that is specific to the reporter, and error.

Metric Invariance Across Time—To test for metric invariance across time, we specified models that fixed factor loadings of the same indicator (i.e., CD, ADHD, or ODD symptoms) to be equal across time, and compared this model with the adjusted initial model in which the loadings were free to vary across time. For example, in the teacher baseline model, the loadings of parent-reported CD symptoms across Grades K, 2, 4, and 5 were fixed to be equal on the general externalizing factors, as were their loadings on the CD-specific factor and the parent-report factor. We tested metric invariance for each type of factor (general, specific, and reporter) in turn. We observed similar results across models. For the teacher baseline model, both the reporter and specific externalizing factors exhibited metric invariance over time, whereas for the parent baseline model, the general and specific externalizing factors exhibited metric invariance over time. On the other hand, for the teacher baseline model, the association of teacher report of ODD with the general externalizing spectrum factor increased over time, with standardized loadings increasing from $.60$ at Grade K to $.88$ at Grade 5, showing evidence of metric noninvariance. Similarly,

in the parent baseline model, teacher report of ODD was noninvariant on the teacher-report factor, increasing similarly over time (from .47 to .86).

For the general externalizing factor, the same pattern of loadings was observed regardless of which reporter served as the “baseline.” For example, in the parent baseline model, over time, teacher-report items contributed more variability ($\lambda = .60$ to $.88$) to the general externalizing factor than did parent-report items ($\lambda = .32$ to $.42$), whereas for the teacher baseline model (when a parent-reported factor was estimated), the opposite was true. The same general pattern held for the parent-reported items. This suggests that, among the current measures, there tended to be more reporter-specific variance in general externalizing behaviors than variance which was common across parent and teacher reports. Finally, the loadings of parent- and teacher-reported symptoms on the ADHD (and parent report of CD and ODD) specific factors were moderate over time.

Scalar Invariance—Next we tested for scalar invariance over time by testing whether the factor intercepts were invariant across time. Table 3 presents the final intercepts for this model. To scale these models, the latent factor variances were fixed to 1, and the latent externalizing, reporter factor means at the first time point and all specific externalizing factor means were fixed to 0 to scale the remaining means, which were freely estimated. Models that freed more latent means were not identified. For the initial scalar invariance model, we fixed all intercepts to be equal across time. Because a model could not be identified when all intercepts were freely estimated, we freed sets of intercepts grouped by reporter/symptom type, comparing each of these models with the fixed intercept model. We tested for scalar invariance of teacher report of ODD, although based on the literature (i.e., Vandenberg & Lance, 2000) we did not expect to find evidence of invariance due to metric noninvariance.

Results were similar across models. In both models, the intercepts of parent-reported ADHD symptoms at Grade 5, and the intercepts of teacher report of ODD symptoms at Grades K and 5 were not equal to those at other time points. All other intercepts were equal within construct and reporter over time.

Factor Means and Variance Invariance—To test for equivalence in the factor variances of the general externalizing factors, we first fixed the variance of the general externalizing factor and reporter factors to 1 across time and freely estimated the remaining factor variances, and then constrained the variances to be equal across time. This freed model did not fit better than the prior model (the final scalar invariance model), suggesting invariance in factor variances over time.

We then constrained the externalizing and reporter factor means to be equal across time, which decreased the fit of the models relative to a model where the Grade K general and parent-report factors were fixed to 0 (models were not identified if one of those means was freely estimated), and the remaining factors across time were freely estimated. Across models, the results were similar. In the parent baseline model, the average level of the general externalizing factor declined over time from a fixed point of .00 at Grade K to $-.30$ at Grade 5, while the level of the teacher-reported externalizing factor was higher at Grades 2 and 5 than at other grades. This pattern was reversed in the teacher baseline model; the

general externalizing was higher at Grades 2 and 5, and the mean of the parent-report factor declined over time. In other words, across 6 years, parents' perception of general externalizing spectrum behaviors declined by .35 *SD*, while teachers' perception of general externalizing behaviors was relatively constant. This final model, with partial metric and scalar invariance, and partial means and variance invariance, fit the data substantially better than the initial model, $BIC = 135$ to 166 . In addition to the improvements in BIC , χ^2 , RMSEA, and CFI exhibited minor changes even with substantial model constraints, and large increases in degrees of freedom.

The Externalizing Spectrum Across Gender

Using the final models above, we then tested whether the structural parameters (factor loadings and intercepts) were invariant across gender. For metric invariance, we first estimated multigroup SEM models constraining all factor loadings across groups, and compared them with models where all loadings were freely estimated across groups (although still constrained across time within groups). We compared models by BIC and examined modification indices and standardized residuals to determine sources of noninvariance.

For both teacher and parent baseline models, there was only minor evidence of metric noninvariance. In the teacher baseline model, teacher report of ODD symptoms at Grade K was more strongly related to the general externalizing factor for males ($\lambda = .70$) than females ($\lambda = .43$), and Grade 5 parent report of ODD symptoms was more strongly related to the general externalizing factor for males ($\lambda = .83$) than females ($\lambda = .75$), all $BIC > 25$. For the parent baseline model, the results were similar: Teacher report of ODD symptoms at Grade K was more strongly related to the teacher-reported externalizing factor for males ($\lambda = .64$) than females ($\lambda = .33$), and Grade 5 parent report of ODD symptoms was more strongly related to the general externalizing factor for males ($\lambda = .85$) than females ($\lambda = .79$), all $BIC > 25$.

In both models, there was substantial scalar noninvariance, such that models where intercepts were freely estimated across gender (but still fixed over time) fit better than a model where intercepts were forced to be equal across groups ($BIC > 43$). Across all constructs, intercepts were higher for males than for females, indicating that at the mean of the latent constructs, males would be expected to be rated as higher than females on CD, ADHD, and ODD symptoms by both teachers and parents.

Figure 2 illustrates the joint impact of metric and scalar noninvariance over time and across gender for teacher report of ODD symptoms. At Grade K, a boy who was 1 *SD* above the mean on the general externalizing spectrum would be expected to have a teacher-reported mean level of ODD symptoms of .61 (on a scale from 0 to 2), whereas the expected level of a girl's teacher-reported ODD symptoms would be .23. By Grade 5, these expected values would be .89 for boys and .64 for girls. Conversely, a teacher-reported ODD symptom average score of .79 at Grade K (the sample mean) would translate into a general externalizing score of 1.63 (i.e., 1 *SD* above the mean) for boys and 5.99 for girls.

Discussion

Multiple prior studies have modeled the covariance of externalizing behaviors in childhood, such as ADHD, ODD, and CD. However, few studies had tested how multiple reporters might see both general and specific covariances among externalizing behaviors, or how their factor structures might differ over time. Using a CT-CM-1 model, our results indicated that parents and teachers shared perspective on some common (approximately 10% to 20%) variance in ADHD, ODD, and CD symptoms between Grades K and 5, while reporter-specific covariation in externalizing problems across childhood explained between 28% and 58% of the variance in symptoms over time. There was stable and specific covariation across time for ADHD symptoms, whereas only parents observed variation in ODD and CD symptoms that was not shared with other externalizing problems. Finally, across time, there was support for metric and partial scalar invariance across time, but there was only partial metric invariance and substantial scalar noninvariance across gender.

The Externalizing Spectrum Across Childhood

Prior research had demonstrated substantial covariation between lifetime diagnoses of ADHD, ODD, and CD during childhood and adolescence (e.g., Essau, 2003), and that, during adolescence, this covariation might be explained by a latent factor (Dick et al., 2005; Tuvblad et al., 2009). Our findings replicated prior work showing that variability in symptoms of ADHD, ODD, and CD during elementary school may be explained by a general liability to externalizing behaviors, but there was also substantial within-disorder variation exhibited over time. Moreover, some variance in this general liability to externalizing behaviors was not shared across parent and teacher report. Using single reporter models or combining across reporters may not fully capture the relations between the externalizing spectrum and its correlates. Finally, unique variance within disorders was observed for parent and teacher report of ADHD, but only for parent report of ODD and CD. This may be because ODD and CD symptoms (such as bullying, fighting, stealing, tantrums, etc.) are observed less often in classroom settings, while ADHD symptoms may be very specific to school settings, and thus more salient and easily distinguished from general externalizing behaviors by teachers.

Measurement Invariance Over Time

Prior research suggests that the rates and presentation of specific externalizing behaviors change with age (Patterson, 1993). We tested whether this means that the actual structure of the externalizing spectrum also changes during middle childhood, and found evidence for metric invariance and scalar invariance across childhood with the exception of teacher report of ODD. In other words, the outward expression of parent and teacher report of ADHD and CD, and parent report of ODD were related to the externalizing spectrum in the same way from kindergarten through fifth grade, and this was true for the general externalizing and reporter-specific externalizing factors. This suggests that the way that symptoms of these disorders are associated with the latent externalizing spectrum is consistent across time. In turn, metric invariance suggests that associations between externalizing spectrum variables and its predictors and outcomes (presuming they are also invariant) are likely to be similar across elementary school, particularly for parent report.

Moreover, we found general support for scalar invariance across development in terms of the externalizing factor intercepts. From kindergarten to Grade 5, at the zero point of the latent factors, the expected parent-reported levels (represented by the factor intercept) of ODD and CD symptoms were identical, and were largely identical for ADHD (except Grade 5 ADHD symptoms), and the same was true for teacher report of CD and ADHD. Moreover, the number of parent-reported symptoms expected at the mean of the externalizing spectrum was highest for ODD, followed by ADHD and then CD symptoms, while the expected number of teacher-reported symptoms at the mean was highest for ADHD, followed by ODD and then CD. This suggests that CD symptoms, in general, may reflect more severe externalizing problems. Prior research has suggested that differences in reports of externalizing behaviors may reflect context-dependent differences in children's behavior (De Los Reyes, Henry, Tolan, & Wakschlag, 2009), as well as differences in the situations in which reporters view the child or in the reporter's normative reference (Achenbach, McConaughy, & Howell, 1987). Teachers frequently play an important role in initial screenings for ADHD symptoms, as difficulties with attention and concentration are often first recognized by teachers (Snider, Busch, & Arrowood, 2003). Additionally, ADHD symptoms are frequently first observed and most troublesome in the classroom where children are required to sustain their attention and refrain from hyperactive or impulsive behaviors (Barkley, 2003). Thus, our findings are consistent with the idea that ADHD symptoms may be particularly salient in school contexts, whereas parents may be particularly sensitive to ODD symptoms as they reflect relational problems, such as arguing and disobedience.

Prior research (e.g., Burns et al., 2006; Guttmanova, Szanyi, & Cali, 2008) found no evidence of scalar noninvariance in the CBCL or other related problem behavior checklists. However, using mean estimates of ADHD, CD, and ODD symptoms, we found metric and scalar noninvariance for teacher report of ODD symptoms. Over time, factor loadings increased, as (generally) did intercepts, indicating that teacher-reported ODD symptoms became more tightly linked with the externalizing spectrum across elementary school, and more ODD symptoms were expected at the mean level of the latent factors in Grade 5 than they did at Grade K. This suggests there may be some differences in the interpretation of overall scores of externalizing symptoms (particularly by teacher report) in terms of how ADHD, ODD, and CD symptoms indicate the broader externalizing spectrum. Although measurement noninvariance does not necessarily mean that prediction will be biased across groups (MacDonald & Paunonen, 2002), future research should investigate what factors across different reporters may influence the predictive validity of externalizing spectrum models, with particular attention to how changes in teacher-reported symptoms may alter the predictive validity of these models.

Although some prior research, confirmed by our findings here, indicated that the general prevalence of externalizing behavior problems declines across childhood (Bongers, Koot, van der Ende, & Verhulst, 2004; Costello et al., 2003), our findings suggest that it is important to attend to the pattern of symptom increases and decreases depending on symptom type and reporter. These differences may reflect an increasing ill-fit of externalizing behaviors within the classroom context, relative to within the home. These findings may also reflect a general population-level decline in externalizing behaviors over

time, making even minor expressions of externalizing-spectrum behaviors appear relatively more severe. Alternately, it may be that some externalizing behaviors (such as fighting) actually *do* increase in severity with age, in terms of both how statistically unusual and how damaging or dangerous they are, but most others decline. Future research should explore the degree to which, at the item level, the symptoms of externalizing behaviors increase or decrease in terms of their relative severity across age to delineate how symptom expression of externalizing behaviors may change across development.

Finally, we found partial metric invariance across gender, and general evidence for scalar noninvariance across gender. For children at the same level of the latent externalizing spectrum, both parents and teachers tended to report higher ADHD, ODD, and CD scores for boys, relative to girls. It may be that externalizing behaviors may be viewed as more normative for boys, and/or that parents and teachers are more sensitized to girls' externalizing behavior during childhood (perhaps because it is more rare). Consistent with our prior findings (Wu et al., 2012) these results suggest that problem behavior measures may be more reflective of overt behavior problems among boys and future research should aim to develop problem behavior measures that are more sensitive to gender differences (e.g., Zoccolillo, 1993).

Limitations of the Current Study

Although the current study has multiple strengths, including the use of a large, community-based sample of children from four regions of the United States that was followed over a long period of time, the use of multiple reporters, and the utilization of advanced measurement modeling techniques, there are also limitations. First, we used a subset of items that were previously identified as representing *DSM-IV* symptoms of ADHD, CD, and ODD, but we did not administer a standardized, structured assessment of symptoms, which may make comparison with other samples or studies more difficult. Indeed, this inconsistency in measurement is common in the literature, with prior studies of the externalizing spectrum using structured or semistructured clinical interviews, computerized diagnostic assessments, or self-reported symptom checklists. As such, it is important to understand the degree to which the measurement structure of the latent spectrum is influenced by measurement mode. Moreover, given that the current sample was largely represented by White and African American children, it may be that the current findings do not generalize to children from other ethnic or racial groups (e.g., Latino, Asian American). This is further complicated by a race by urban/rural confound in the Fast Track Project, as virtually all the Black participants lived in urban areas. Given that we had no a priori hypotheses regarding race and urban rural status, and a detailed examination of these effects would have further divided the current sample (by age, gender, and race/urban/rural status), the current sample was underpowered to provide a full test of these effects. Although we weighted the current models by each child's probability of selection, it may be that incorporating the high-risk children reflected a population that exhibits greater overlap between ADHD, ODD, and CD. There has been less consideration of whether the externalizing spectrum has a different structure in high versus low risk or representative samples. Thus, it is important to understand the structure of the externalizing spectrum in high-risk samples, in part because these are the children who may be more likely to be seen

in clinical settings. Additionally, because the Fast Track study did not begin until kindergarten, and the CBCL was not administered during high school, we did not examine the structure of the externalizing spectrum during early childhood or during middle to late adolescence. Future research should attempt to determine how early externalizing spectrum behaviors begin to emerge, and how they progress across adolescence. In addition, our symptom ratings were not ideal, which may have biased the current models. Specifically, the symptom counts (i.e., *DSM-IV*-derived scales from the CBCL) did not include all symptoms of each disorder, which may have underestimated the prevalence of symptoms in the sample. Finally, given some prior research suggesting that emotion dysregulation is central to ADHD (Barkley, 2015), it may be that the current models overestimate the associations between ADHD symptoms and the externalizing spectrum. Future research should work to better clarify the role of ADHD in the internalizing/externalizing spectra.

Summary and Future Directions

In sum, measurement models of the externalizing spectrum provided evidence that an underlying risk factor for covariation among symptoms of ADHD, ODD, and CD is observable and stable (but with changing symptom levels) across much of childhood. There was evidence for continuity and stability across development and across reporters. Finally, future research should aim to identify contextual factors (such as parenting) that strengthen or weaken the associations between specific disorders (such as ADHD) and the externalizing spectrum, in order to improve targeted prevention and intervention efforts. In short, the externalizing spectrum may be a useful heuristic for describing covariance among many externalizing symptoms, but it does not explain fully why any particular externalizing disorder may or may not develop.

Acknowledgments

This work used data from the Fast Track project (for additional information concerning Fast Track, see <http://www.fasttrackproject.org>). We are grateful for the collaboration of the Durham Public Schools, the Metropolitan Nashville Public Schools, the Bellefonte Area Schools, the Tyrone Area Schools, the Mifflin County Schools, the Highline Public Schools, and the Seattle Public Schools. We appreciate the hard work and dedication of the many staff members who implemented the project, collected the evaluation data, and assisted with data management and analyses. We thank Robert Krueger for his consultation on this project.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by National Institute of Mental Health (NIMH) Grants R18 MH48043, R18 MH50951, R18 MH50952, R18 MH50953, K05MH00797, and K05MH01027; National Institute on Drug Abuse (NIDA) Grants DA016903, K05DA15226, and P30DA023026; and Department of Education Grant S184U30002. The Center for Substance Abuse Prevention also provided support through a memorandum of agreement with the NIMH. Additional support for the preparation of this work was provided by a LEEF B.C. Leadership Chair award, Child & Family Research Institute Investigator Salary and Investigator Establishment Awards, and a Canada Foundation for Innovation award to Robert J. McMahon.

References

- Achenbach TM. The Child Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology*. 1978; 46:478–488. [PubMed: 670491]
- Achenbach TM. Manual for the Child Behavior Checklist/4–18 and 1991 Profile. Burlington: University of Vermont Department of Psychiatry; 1991a.

- Achenbach TM. Manual for the Teacher's Report Form and 1991 Profile. Burlington: University of Vermont Department of Psychiatry; 1991b.
- Achenbach TM, Dumenci L, Rescorla LA. DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology*. 2003; 32:328–340. [PubMed: 12881022]
- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213. [PubMed: 3562706]
- Allison PD. Missing data. Thousand Oaks, CA: Sage; 2002.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000. text rev
- Angold A, Costello EJ, Erkanli A. Comorbidity. *Journal of Child Psychology and Psychiatry*. 1999; 40:57–87. [PubMed: 10102726]
- Barkley RA. Attention-deficit/hyperactivity disorder. In: Mash EJ, Barkley RA, editors *Child psychopathology*. 2. New York, NY: Guilford Press; 2003. 75–143.
- Barkley RA. Emotional dysregulation is a core component of ADHD. In: Barkley RA, editor *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment*. 4. New York, NY: Guilford Press; 2015. 81–115.
- Beauchaine TP, Hinshaw SP, Pang KL. Comorbidity of attention-deficit/hyperactivity disorder and early-onset conduct disorder: Biological, environmental, and developmental mechanisms. *Clinical Psychology: Science and Practice*. 2010; 17:327–336.
- Beauchaine TP, McNulty T. Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology. *Development and Psychopathology*. 2013; 25:1505–1528. DOI: 10.1017/S0954579413000746 [PubMed: 24342853]
- Beesdo-Baum K, Höfler M, Gloster AT, Klotsche J, Lieb R, Beauducel A, Wittchen HU. The structure of common mental disorders: A replication study in a community sample of adolescents and young adults. *International Journal of Methods in Psychiatric Research*. 2009; 18:204–220. [PubMed: 20024895]
- Bezdjian S, Krueger RF, Derringer J, Malone S, McGue M, Iacono WG. The structure of DSM-IV ADHD, ODD, and CD criteria in adolescent boys: A hierarchical approach. *Psychiatry Research*. 2011; 188:411–421. [PubMed: 21470694]
- Biederman JJ, Petty CR, Dolan CC, Hughes SS, Mick EE, Monuteaux MC, Faraone SV. The long-term longitudinal course of oppositional defiant disorder and conduct disorder in ADHD boys: Findings from a controlled 10-year prospective longitudinal follow-up study. *Psychological Medicine*. 2008; 38:1027–1036. [PubMed: 18205967]
- Bongers IL, Koot HM, van der Ende J, Verhulst FC. Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*. 2004; 75:1523–1537. [PubMed: 15369529]
- Bornovalova MA, Hicks BM, Iacono WG, McGue M. Familial transmission and heritability of childhood disruptive disorders. *American Journal of Psychiatry*. 2010; 167:1066–1074. [PubMed: 20634367]
- Burns GL, de Moura MA, Beauchaine TP, McBurnett K. Bifactor latent structure of ADHD/ODD symptoms: Predictions of dual-pathway/trait-impulsivity etiological models of ADHD. *Journal of Child Psychology and Psychiatry*. 2014; 55:393–401. [PubMed: 24795957]
- Burns GL, Walsh JA, Gomez R, Hafetz N. Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*. 2006; 18:452–457. [PubMed: 17154767]
- Burt SA. Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. *Psychological Bulletin*. 2009; 135:608–637. [PubMed: 19586164]
- Conduct Problems Prevention Research Group. A developmental and clinical model for the prevention of conduct disorders: The FAST Track program. *Development and Psychopathology*. 1992; 4:509–527.

- Conduct Problems Prevention Research Group. Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology*. 1999; 67:631–647. [PubMed: 10535230]
- Conduct Problems Prevention Research Group. Merging universal and indicated prevention programs: The Fast Track model. *Addictive Behaviors*. 2000; 25:913–927. [PubMed: 11125779]
- Conduct Problems Prevention Research Group. Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology*. 2002; 30:19–35. [PubMed: 11930969]
- Connor DF, Steeber J, McBurnett K. A review of attention-deficit/hyperactivity disorder complicated by symptoms of oppositional defiant disorder or conduct disorder. *Journal of Developmental and Behavioral Pediatrics*. 2010; 31:427–440. [PubMed: 20535081]
- Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*. 2003; 60:837–844. [PubMed: 12912767]
- De Los Reyes A, Henry DB, Tolan PH, Wakschlag LS. Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*. 2009; 37:637–652. [PubMed: 19247829]
- Diamantopoulou S, Verhulst FC, van der Ende J. The parallel development of ODD and CD symptoms from early childhood to adolescence. *European Child & Adolescent Psychiatry*. 2011; 20:301–309. [PubMed: 21499848]
- Dick DM, Viken RJ, Kaprio J, Pulkkinen L, Rose RJ. Understanding the covariation among childhood externalizing symptoms: Genetic and environmental influences on conduct disorder, attention deficit hyperactivity disorder, and oppositional defiant disorder symptoms. *Journal of Abnormal Child Psychology*. 2005; 33:219–229. [PubMed: 15839499]
- Dodge KA, Malone PS, Lansford JE, Miller S, Pettit GS, Bates JE. A dynamic cascade model of the development of substance-use onset. *Monographs of the Society for Research in Child Development*. 2009; 74(3):vii–119. DOI: 10.1111/j.1540-5834.2009.00528.x
- Eid M, Lischetzke T, Nussbeck FW, Trierweiler LI. Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-CM-1 model. *Psychological Methods*. 2003; 8:38–60. [PubMed: 12741672]
- Eid M, Nussbeck FW, Geiser C, Cole DA, Gollwitzer M, Lischetzke T. Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*. 2008; 13:230–253. [PubMed: 18778153]
- Essau CA. *Conduct and oppositional defiant disorders*. London, England: Lawrence Erlbaum; 2003.
- Guttmannova K, Szanyi JM, Cali PW. Internalizing and externalizing behavior problem scores: Cross-ethnic and longitudinal measurement invariance of the Behavior Problem Index. *Educational and Psychological Measurement*. 2008; 68:676–694.
- Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Hudziak JJ, Copeland W, Stanger C, Wadsworth M. Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: A receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry*. 2004; 45:1299–1307. [PubMed: 15335349]
- Jackson DL, Gillaspie JA Jr, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*. 2009; 14:6–23. [PubMed: 19271845]
- Jones D, Dodge KA, Foster EM, Nix R. the Conduct Problems Prevention Research Group. Early identification of children at risk for costly mental health service use. *Prevention Science*. 2002; 3:247–256. [PubMed: 12458763]
- Kenny DA, McCoach DB. Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*. 2003; 10:333–351.
- Knopik VS, Heath AC, Bucholz KK, Madden PAF, Waldron M. Genetic and environmental influences on externalizing behavior and alcohol problems in adolescence: A female twin study. *Pharmacology Biochemistry and Behavior*. 2009; 93:313–321.

- Krol NP, De Bruyn EE, Coolen JC, van Aarle EJ. From CBCL to DSM: A comparison of two methods to screen for DSM-IV diagnoses using CBCL data. *Journal of Clinical Child and Adolescent Psychology*. 2006; 35:127–135. [PubMed: 16390308]
- Krueger RF, Hicks BM, Patrick CJ, Carlson SR, Iacono WG, McGue M. Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*. 2002; 111:411–424. [PubMed: 12150417]
- Lahey BB, Van Hulle CA, Rathouz PJ, Rodgers J, D'Onofrio BM, Waldman ID. Are oppositional-defiant and hyperactive-inattentive symptoms developmental precursors to conduct problems in late childhood? Genetic and environmental links. *Journal of Abnormal Child Psychology*. 2009; 37:45–58. [PubMed: 18648930]
- Little RJA, Rubin DB. *Statistical analysis with missing data*. 2. New York, NY: Wiley-Interscience; 2002.
- Loeber R, Burke JD. Developmental pathways in juvenile externalizing and internalizing problems. *Journal of Research on Adolescence*. 2011; 21:34–46. [PubMed: 22468115]
- Loeber R, Burke JD, Lahey BB, Winters A, Zera M. Oppositional defiant and conduct disorder: A review of the past ten years, part 1. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39:1468–1484. [PubMed: 11128323]
- Loeber R, Burke JD, Pardini DA. Development and etiology of disruptive and delinquent behavior. *Annual Review of Clinical Psychology*. 2009; 5:291–310.
- Lorber MF, Del Vecchio T, Slep AM. The emergence and evolution of infant externalizing behavior. *Development and Psychopathology*. 2014; 27:663–680. [PubMed: 25212988]
- MacDonald P, Paunonen SV. A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*. 2002; 62:921–943.
- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341.
- Martel MM, Pierce L, Nigg JT, Jester JM, Adams K, Puttler LI, Zucker RA. Temperament pathways to childhood disruptive behavior and adolescent substance abuse: Testing a cascade model. *Journal of Abnormal Child Psychology*. 2009; 37:363–373. [PubMed: 18787942]
- Meredith W. Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*. 1993; 58:525–543.
- Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*. 2009; 11:7–20. [PubMed: 19432384]
- Millsap RE. Invariance in measurement and prediction: Their relationship in the single factor case. *Psychological Methods*. 1997; 3:248–260.
- Molina BSG. Delinquency and substance use in ADHD: Adolescent and young adult outcomes in developmental context. In: Jensen PD, Cooper JR, editors *Attention deficit hyperactivity disorder: State of the science & best practices*. Vol. 2. New York, NY: Civic Research Institute; 2011. 19–19-12.
- Muthén L, Muthén B. *Mplus user's guide: Statistical analysis with latent variables*. 6. Los Angeles, CA: Muthén & Muthén; 2010.
- Nadder TS, Rutter M, Silberg JL, Maes HH, Eaves LJ. Genetic effects on the variation and covariation of attention deficit-hyperactivity disorder (ADHD) and oppositional-defiant disorder/conduct disorder (ODD/CD) symptomologies across informant and occasion of measurement. *Psychological Medicine*. 2002; 32:39–53. [PubMed: 11883729]
- Olson SL, Sameroff AJ, Lansford JE, Sexton H, DavisKean P, Bates JE, Dodge KA. Deconstructing the externalizing spectrum: Growth patterns of overt aggression, covert aggression, oppositional behavior, impulsivity/inattention, and emotion dysregulation between school entry and early adolescence. *Development and Psychopathology*. 2013; 25:817–842. [PubMed: 23880394]
- Patterson GR. Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology*. 1993; 61:911–919. [PubMed: 8113492]

- Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*. 1993; 114:552–566. [PubMed: 8272470]
- Reitz E, Dekovic M, Meijer AM. The structure and stability of externalizing and internalizing problem behavior during early adolescence. *Journal of Youth and Adolescence*. 2005; 34:577–588.
- Silberg J, Rutter M, Meyer J, Maes H, Hewitt J, Simonoff E, Eaves L. Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. *Journal of Child Psychology and Psychiatry*. 1996; 37:803–816. [PubMed: 8923223]
- Smith GT. On construct validity: issues of method and measurement. *Psychological assessment*. 2005; 17(4):396. [PubMed: 16393005]
- Snider VE, Busch T, Arrowood L. Teacher knowledge of stimulant medication and ADHD. *Remedial and Special Education*. 2003; 24:46–56.
- Tackett JL, Daoud SL, De Bolle M, Burt SA. Is relational aggression part of the externalizing spectrum? A bifactor model of youth antisocial behavior. *Aggressive Behavior*. 2013; 39:149–159. DOI: 10.1002/ab.21466 [PubMed: 23386551]
- Tackett JL, Herzhoff K, Reardon KW, De Clercq B, Sharp C. The externalizing spectrum in youth: Incorporating personality pathology. *Journal of Adolescence*. 2014; 37:659–668. [PubMed: 24263222]
- Thapar A, Harrington R, McGuffin P. Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. *British Journal of Psychiatry*. 2001; 179:224–229. [PubMed: 11532799]
- Tuvblad C, Zheng M, Raine A, Baker LA. A common genetic factor explains the covariation among ADHD, ODD, and CD symptoms in 9–10 year old boys and girls. *Journal of Abnormal Child Psychology*. 2009; 37:153–167. [PubMed: 19015975]
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–69.
- Waschbusch DA. A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin*. 2002; 128:118–150. [PubMed: 11843545]
- Werthamer-Larsson L, Kellam SG, Wheeler L. Effects of first grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology*. 1991; 19:585–560. [PubMed: 1755437]
- Wittchen HU, Beesdo-Baum K, Gloster AT, Höfler M, Klotsche J, Lieb R, Kessler RC. The structure of mental disorders re-examined: Is it developmentally stable and robust against additions? *International Journal of Methods in Psychiatric Research*. 2009; 18:189–203. [PubMed: 20033884]
- Wu J, King KM, Witkiewitz K, McMahon RJ, the Conduct Problems Prevention Research Group. Item analysis and differential item functioning of a brief conduct problem screen. *Psychological Assessment*. 2012; 24:444–454. [PubMed: 22040514]
- Young SE, Friedman NP, Miyake A, Willcutt EG, Corley RP, Haberstick BC, Hewitt JK. Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*. 2009; 118:117–130. [PubMed: 19222319]
- Zoccolillo M. Gender and the development of conduct disorder. *Development and Psychopathology*. 1993; 5:65–78.

Figure 1.

An illustration of the final CT-CM-1 model of the common-specific model of the externalizing spectrum.

Note. CD = conduct disorder; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder; CT-CM-1 = correlated-trait, multimethod minus one. In a parallel model, we modeled the parent-report factors to replace the teacher-report factors.

Figure 2.

The impact of metric and scalar noninvariance on the measurement of the general externalizing spectrum at (a) Grade K and (b) Grade 5.

Note. An illustration of the impact of noninvariance. Vertical axes are scaled from the minimum to +1 *SD* above the mean of ODD symptoms; horizontal axes are scaled from -2 to +2 *SD* of general externalizing. The slopes for boys and girls show the expected level of teacher-report ODD symptoms for a given level of the latent general externalizing spectrum.

Table 1

Descriptive Statistics of All Study Variables (Includes Normative and High-Risk Control Individuals Combined)

Variable	Grade K (n = 748)	Grade 2 (n = 687)	Grade 4 (n = 669)	Grade 5 (n = 647)
<i>Childhood indicators</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>	<i>M (SE)</i>
ADHD symptoms (Parent)	1.48 (0.04)	1.63 (0.05)	1.45 (0.05)	1.36 (0.05)
ADHD symptoms (Teacher)	2.03 (0.04)	2.08 (0.04)	1.63 (0.05)	1.50 (0.05)
ODD symptoms (Parent)	3.35 (0.06)	3.17 (0.06)	2.88 (0.06)	2.80 (0.07)
ODD symptoms (Teacher)	3.87 (0.12)	4.16 (0.13)	3.34 (0.12)	3.39 (0.12)
CD symptoms (Parent)	2.14 (0.08)	1.97 (0.08)	1.72 (0.09)	1.67 (0.08)
CD symptoms (Teacher)	2.57 (0.10)	2.73 (0.12)	1.92 (0.09)	2.00 (0.10)
	α	α	α	α
ADHD symptoms (parent)	.63	.70	.72	.73
ODD symptoms (parent)	.68	.70	.72	.76
CD symptoms (parent)	.77	.77	.80	.78
ADHD symptoms (teacher)	.92	.91	.92	.92
ODD symptoms (teacher)	.90	.88	.91	.91
CD symptoms (teacher)	.85	.85	.86	.85

Note. CD = conduct disorder; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder; α = Cronbach's alpha.

Table 2

Model Fit Statistics From the Final Models

	Model fit statistics parent baseline					Model fit statistics teacher baseline				
	χ^2	<i>df</i>	CFI	RMSEA	BIC	χ^2	<i>df</i>	CFI	RMSEA	BIC
Baseline model	710.28	204	.908	.057	1,040	774.02	204	.897	.061	1,131
Adjusted baseline model	711.01	212	.909	.056	999	772.71	212	.898	.059	1,089
Metric invariance fixed general externalizing loadings	813.09	230	.894	.058	1,059	780.316	230	.900	.056	1,005
Metric invariance, fixed reporter loadings	712.12	221	.911	.054	955	872.244	221	.882	.063	1,196
Metric invariance fixed specific externalizing loadings	750.53	224	.904	.056	1,005	801.872	224	.895	.059	1,085
Metric invariance final	761.26	248	.907	.052	907	819.51	248	.90	.055	992
Scalar invariance initial	821.67	260	.898	.054	918	880.91	260	.89	.060	1,013
Scalar invariance final	777.01	257	.906	.052	864	840.41	257	.89	.055	965
Factor variances freed	778.01	251	.904	.053	892	840.82	251	.893	.056	988
Factor means fixed	832.51	261	.896	.054	931	905.04	262	.883	.057	1,037
Factor means final	777.02	257	.906	.052	864	840.41	257	.89	.055	965

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; BIC = Bayesian information criteria; *df* = degrees of freedom. The best fitting model is highlighted in bold; it was not identical to the final scalar invariance model. Where invariance held, the initial and final model fit was identical. "Initial" model refers to the initial test of invariance, where parameters were fixed over time. "Final" model refers to the final invariance model, with some parameters (see text) freed over time.

Table 3

Standardized Factor Loadings, Intercepts, and R^2 Values for the Teacher and Parent Baseline Models

Teacher baseline Model	Parent-reported items					Teacher-reported items				
	$F_{\text{Externalizing}}$	F_{Disorder}	F_{Parent}	Male intercept	Female intercept	R^2	$F_{\text{Externalizing}}$	F_{Disorder}	F_{Parent}	R^2
ADHD Grade K	.43	.47	.55	1.27	1.08	.71	.73	.29	—	.63
ADHD Grade 2	.43	.47	.55	1.27	1.04	.73	.73	.31	—	.54
ADHD Grade 4	.43	.47	.55	1.27	1.08	.73	.73	.27	—	.69
ADHD Grade 5	.43	.47	.55	1.30	1.13	.78	.73	.31	—	.61
ODD Grade K	.32	.36	.74	1.67	1.33	.78	.69 (.51)	—	—	.36
ODD Grade 2	.32	.36	.74	1.71	1.37	.83	.66	—	—	.41
ODD Grade 4	.32	.36	.74	1.66	1.39	.82	.67	—	—	.45
ODD Grade 5	.32 (.25)	.36	.74	1.59	1.36	.82	.88	—	—	.78
CD Grade K	.34	.35	.62	0.83	0.59	.62	.77	—	—	.59
CD Grade 2	.34	.35	.62	0.88	0.61	.68	.77	—	—	.53
CD Grade 4	.34	.35	.62	0.95	0.62	.75	.77	—	—	.66
CD Grade 5	.34	.35	.62	0.96	0.61	.78	.77	—	—	.53

Parent baseline Model	Parent-reported items					Teacher-reported items				
	$F_{\text{Externalizing}}$	F_{Disorder}	F_{Teacher}	Male intercept	Female intercept	R^2	$F_{\text{Externalizing}}$	F_{Disorder}	F_{Teacher}	R^2
ADHD Grade K	.72	.46	—	1.27	1.32	.72	.27	.33	.63	.60
ADHD Grade 2	.72	.46	—	1.33	1.42	.74	.27	.33	.63	.51
ADHD Grade 4	.72	.46	—	1.31	1.33	.75	.27	.33	.63	.66
ADHD Grade 5	.72	.46	—	1.30	1.27	.80	.27	.33	.63	.67
ODD Grade K	.78	.37	—	1.46	1.50	.75	.30	—	.61 (.38)	.32
ODD Grade 2	.78	.37	—	1.52	1.55	.80	.30	—	.58	.37
ODD Grade 4	.78	.37	—	1.58	1.57	.79	.30	—	.57	.37
ODD Grade 5	.78 (.73)	.37	—	1.39	1.54	.79	.30	—	.86	.78
CD Grade K	.72	.34	—	0.72	0.73	.63	.28	—	.73	.60
CD Grade 2	.72	.34	—	0.76	0.77	.68	.28	—	.73	.52
CD Grade 4	.72	.34	—	0.81	0.90	.74	.28	—	.73	.64
CD Grade 5	.72	.34	—	0.82	0.89	.75	.28	—	.73	.51

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Note. $F_{\text{Externalizing}}$ = factor loadings for the externalizing factor; F_{Disorder} = factor loadings for the disorder-specific factor; F_{Parent} = factor loadings for the parent-report factor; F_{Teacher} = factor loadings for the teacher-report factor; CD = conduct disorder; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder. Parameters in boldface exhibited noninvariance across time, loadings in parentheses exhibited invariance across gender. All factor loadings and intercepts are standardized. Because parameters that are fixed to be equal can differ in standardized form because of differences in item variances over time, we display standardized estimates from Grade K *except* where parameters were not fixed to be equal over time.

Table 4

Latent Factor Correlations From the Final Models

Reporter-specific externalizing	Grade K	Grade 2	Grade 4	Grade 5
Grade K		.63	.66	.70
Grade 2	.64		.79	.78
Grade 4	.49	.58		1.00
Grade 5	.48	.57	.72	
General externalizing	Grade K	Grade 2	Grade 4	Grade 5
Grade K		.72	.62	.67
Grade 2	.77		.69	.67
Grade 4	.75	.88		.83
Grade 5	.81	.83	1.00	

Note. Correlations above the diagonal are from the parent baseline model, correlations below the diagonal are from the teacher baseline model.