Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_______________________________ _______________

Ryan Hackett Date

Disruptive Behavior Disorders, Specific Parenting Practices, and Broad Dimensions of Parenting: Factor Structure, and Measurement, Structural, and Prediction Invariance

By

Ryan C. Hackett
Master of Arts

Clinical Psychology

Irwin Waldman, Ph.D.
Advisor

Robyn Fivush, Ph.D.
Committee Member

Scott O. Lilienfeld, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Disruptive Behavior Disorders, Specific Parenting Practices, and Broad Dimensions of Parenting: Factor Structure and Measurement, Structural, and Prediction Invariance

By

Ryan C. Hackett
A.B., Harvard University

Advisor: Irwin Waldman, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University in partial fulfillment of
the requirements for the degree of Master of Arts
in Psychology
2016

Abstract

Disruptive Behavior Disorders, Specific Parenting Practices, and Broad Dimensions of Parenting: Factor Structure and Measurement, Structural, and Prediction Invariance
By Ryan C. Hackett

When questionnaire scales are used in an analysis, it is assumed that their psychometric properties are invariant across key grouping variables (Millsap, 2011). If this assumption, known as measurement invariance, is violated, phenotypic group differences can be confounded by measurement differences, and relations with external variables can be estimated inaccurately (Millsap, 2011). Despite the importance of measurement invariance, it is rarely examined in the parenting and psychopathology literature. Prior research has suggested small-to-moderate relations between parenting and youth antisocial behavior in samples that varied widely in sex and age, with little examination of sex or age invariance (e.g., Hoeve et al., 2009). This study evaluated the factor structure, and sex and age measurement, structural, and prediction invariance of the Modified Child-Rearing Practices Report (M-CRPR, Rickel & Biasatti, 1982), which measures parental responsiveness and psychological control, Supervision/Involvement scale (S/I, Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998), which measures parental involvement and oversight, and Emory Diagnostic Rating Scale (EDRS, Waldman et al., 1998), which measures DSM-IV symptoms of oppositional defiant disorder (ODD) and conduct disorder (CD). The M-CRPR, S/I, and EDRS ODD scales demonstrated near complete measurement and structural invariance in separate sex and age multigroup invariance analyses. The EDRS CD scale demonstrated item threshold invariance in multiple-indicators-multiple-causes (MIMIC) analyses (Jöreskog & Goldberger, 1975). All relations between parenting and disruptive behavior were invariant and small-to-moderate in magnitude. These findings add confidence to prior research relying on these measures to examine parenting influences on child disruptive behavior.

Disruptive Behavior Disorders, Specific Parenting Practices, and Broad Dimensions of Parenting: Factor Structure and Measurement, Structural, and Prediction Invariance

By

Ryan C. Hackett
A.B., Harvard University

Advisor: Irwin Waldman, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University in partial fulfillment of
the requirements for the degree of Master of Arts
in Psychology
2016

Acknowledgements

I would like to thank Drs. Waldman, Fivush, and Lilienfeld for their encouragement and guidance during this challenging process. I would also like to thank the Emory Department of Psychology and the Laney Graduate School for the excellent educational opportunities they have provided. Finally, I would like to thank my family for the sacrifices they make every day so that I can pursue higher education, and for their unfailing love and affection.

Table of Contents

Disruptive Behavior Disorders, Specific Parenting Practices, and Broad Dimensions of Parenting: Factor Structure, and Measurement, Structural, and Prediction Invariance

The disruptive behavior disorders (DBDs), including oppositional defiant disorder (ODD) and conduct disorder (CD), are among the most common causes for referral to youth mental health services (Kazdin, 1987). ODD is characterized by a persistent pattern of irritable, noncompliant, antagonistic, and vindictive interpersonal behavior that lasts at least 6 months and causes significant distress and/or impairment (American Psychiatric Association, 2013). CD is a persistent pattern of behavior that violates the rights of others and/or major age-appropriate norms and rules that occurs consistently within a 12-month period and causes significant impairment (American Psychiatric Association, 2013). In Medicaid costs alone, average expenditures for youth with ODD and CD may be 25-34 times that for youth with no psychiatric disorder (Mandell, Guevara, Rostain, & Hadley, 2003). These costs may continue across the life course. For example, in a large nationally representative epidemiological sample, 94% of individuals with a lifetime diagnosis of ODD were estimated to suffer at least one additional lifetime mood (46%), anxiety (62%), impulse-control (68%), and substance use disorder (47%), and 64% of individuals with a lifetime diagnosis of ODD were estimated to suffer three or more disorders, with ODD often emerging first (Nock, Kazdin, Hiripi, & Kessler, 2007). Among common youth mental health disorders, CD is associated with functional impairments that are more chronic, global, severe, and prone to relapse (Lambert, Wahler, Andrade, & Bickman, 2001), and as adults, youth with CD suffer increased risk of criminal behavior (Fergusson, Horwood, & Ridder, 2005), antisocial personality disorder (Burke, Waldman, & Lahey, 2010b) and suicide (Bridge, Goldstein, & Brent, 2006). Without successful intervention, a single youth with severe and persistent antisocial behavior who goes on to become a career criminal could cost society as much as 5.8 million dollars by the time they reach age 26[1] (Cohen & Piquero, 2009).

Given the prevalence and consequences of these disorders, much research has been devoted to understanding their etiology, prevention, and remediation. A particularly plausible mechanism in the etiology of ODD and CD is parenting (Kazdin, 2008). Key constructs include specific parenting practices, such as monitoring, discipline, and involvement (e.g., Hirschi, 1969; West & Farrington, 1973; McCord, 1979; Loeber & Stouthamer-Loeber, 1986; Moffitt, 1993; Sampson & Laub, 1994; Dishion & Patterson, 2006) and broad dimensions of parenting, such as responsiveness, psychological control, and behavioral control (e.g., Baumrind, 1971; Maccoby & Martin, 1983; Barber, Olsen, & Shagle, 1994; Hoeve et al., 2009). Overall, meta-analyses of both correlational (e.g., Loeber & Stouthamer-Loeber, 1986; Hoeve et al., 2009) and intervention studies (e.g., McCart, Priester, Davies, & Azen, 2006; Kaminski, Valle, Filene, & Boyle, 2008) converge in estimating small-to-moderate effects of parenting on ODD, CD, and related constructs. The tiny subset of randomized controlled trials examining mediation of treatment effects by parenting have provided some corroboration of these more general findings (e.g., Forgatch, Patterson, DeGarmo, & Beldavs 2009), although results have been mixed, sample sizes have been small, and these studies have remained silent on parenting as an original cause (Forehand, Lafko, Parent, & Burt, 2014).

Even more rare than rigorous experimental tests of parenting as an original and/or maintaining condition of antisocial behavior are studies testing the equivalence of the

[1] Costs to "society" operationalized as costs to victims, the criminal justice system, the penal system, and opportunity costs suffered by offenders

psychometric properties of the surveys used to assess parenting and disruptive behavior across important sub-groups or with repeated use over time. Neglect of measurement invariance is not exclusive to the parenting and psychopathology literature (Lubke, Dolan, & Neale, 2004; Borsboom, 2006; Millsap, 2007; Chen, 2008), but this assumption must be met for valid between-groups and cross-time comparisons, as well as the valid use of heterogeneous samples (Meredith, 1993; Millsap, 2011). A review of the parenting and psychopathology literature easily identifies instances where this assumption may be violated (e.g., Butler, 2013).

The primary goal of the current study is to examine the assumption of measurement invariance across sex and age for two widely used assessments of self-reported parenting, the Modified Child-Rearing Practices Report (M-CRPR) (Rickel & Biasatti, 1982) and the Supervision/Involvement Scale (S/I) (Loeber et al., 1998), as well as two parent-report DSM-IV ODD and CD symptom rating scales from the Emory Diagnostic Rating Scale (EDRS, Waldman et al., 1998). The M-CRPR measures attitudes related to parental responsiveness and psychological control (Rickel & Biasatti, 1982; Deković, Janssens, & Gerris, 1991; Mason, Cauce, Gonzales, & Hiraga, 1996). The S/I scale measures parental oversight and positive involvement (Loeber et al., 1998). Although each of these four measures have been used to examine a variety of questions related to the etiology and course of youth disruptive behavior, none have been examined for measurement invariance across sex and age, and the internal structures of the S/I and EDRS CD scales are also unknown.

Measurement, Structural, and Prediction Invariance

For the assumption of measurement invariance to be minimally satisfied in questionnaire measures using ordinal items, three conditions must be met (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). First, the number of latent constructs (factors) assumed to explain the shared variance among scale items must be equivalent across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). This is known as configural invariance (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Second, the magnitude and direction of the relations between items and factors (factor loadings) must be equivalent across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). This is known as weak or metric invariance (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Third, the point(s) on the continuous normal latent trait distribution assumed to underlie the manifest (observed) responses to each item at which a subject's response changes from one manifest value to another (item threshold(s)) must be equivalent across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). This is known as strong or scalar invariance (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Satisfaction of these three conditions suggests that the constructs measured, the meaning of the items in relation to the constructs measured, and the point at which a person changes their response from one value to another (e.g., from "rarely" to "sometimes") on each item are equivalent across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). These three conditions satisfy the requirement that the questionnaire is "measuring the same thing" (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Violation of any of these conditions can confound between-groups and cross-time comparisons on the constructs underlying a measure, can create biases in parameter estimates for the measurement models of the target constructs, and can create biases in parameter estimates for the relations between the target constructs and other variables (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011).

When researchers are interested in calculating manifest scale scores using raw data, a fourth condition must be met (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). This condition requires that the amount of information in each item that is not explained by the factors (item residual variance) must be equivalent across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). This is known as strict invariance (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Satisfaction of this condition suggests that each item contains the same amount of unique variance (including measurement error) across groups and repeated assessments (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011).

Structural invariance - the equivalence of the variances, co-variances, and means of factors across groups and repeated assessments – is not a requirement for the valid use of a measure, but can still have important implications (Meredith, 1993; Millsap, 2011). This is particularly true for multidimensional measures in which the magnitude and direction of factor correlations are of theoretical interest (Meredith, 1993; Millsap, 2011). Even with measurement and structural invariance established, the magnitude and direction of the relations between the target constructs of a measure and external variables cannot be assumed to be equivalent across groups and repeated assessments (Millsap, 1995). This assumption, prediction invariance, must be tested empirically (Millsap, 1995). If it is violated, then parameter estimates of path coefficients between target constructs and external variables will be biased in samples collapsing across multiple groups/time points (Millsap, 1995).

Tools for Testing Factor Structure and Invariance

Confirmatory factor analysis (CFA) is founded in the independent cluster model, which assumes that each item in a scale has precisely one nonzero factor loading, and therefore all items are infallible measures of single constructs (Asparouhov & Muthén, 2009; Marsh, Morin, Parker, & Kaur, 2014). In line with this assumption, CFA requires investigators to specify the number of factors and the locations of all nonzero factor loadings a priori, and allows hypothesized measurement models to be tested against plausible alternatives (Jöreskog, 1969). In theory, these restrictions encourage researchers to design measures with a simple structure, and to pay heed to prior knowledge in the formalization and testing of measurement hypotheses (Asparouhov & Muthén, 2009).

Unfortunately, CFA models often fail to achieve adequate model fit due to the fallible nature of items as indicators of single constructs, requiring multiple rounds of post-hoc model modifications that may fail to replicate in later samples (Browne, 2001; Asparouhov & Muthén, 2009; Marsh et al., 2014). In addition, the restriction of naturally occurring item cross-loadings creates an upward bias in the magnitude of factor correlations (Asparouhov & Muthén, 2009; Marsh et al., 2014). This upward bias leads to artificial reductions in the discriminant validity of factors (Asparouhov & Muthén, 2009; Marsh et al., 2014). This bias in factor correlations leads to non-essential multicollinearity in path analyses (Asparouhov & Muthén, 2009; Marsh et al., 2014).

These issues have never arisen in exploratory factor analysis (EFA), which provides empirically derived measurement models using factor rotation and places no restrictions on the number of nonzero loadings per item (Browne, 2001; Asparouhov & Muthén, 2009; Marsh et al., 2014). This free estimation of item cross-loadings can be justified based on either substantive or methodological grounds (Browne, 2001; Asparouhov & Muthén, 2009; Marsh et al., 2014).

Recently, a more generalized framework has emerged combining the advantages of CFA and EFA called Exploratory Structural Equation Modeling (ESEM, Asparouhov & Muthén, 2009). This method incorporates EFA factors into the CFA/SEM framework, offering the option of empirically derived measurement models with freely estimated item cross-loadings that can be used with nearly the same flexibility as in CFA (Asparouhov & Muthén, 2009; Marsh et al., 2014). This allows researchers to rely on more “realistic” latent variable measurement models with better discriminant validity in path analyses and tests of invariance (Asparouhov & Muthén, 2009; Marsh et al., 2014).

The advantages of ESEM over CFA must be balanced against the number of additional parameters estimated when item cross-loadings are unrestricted. In addition, like EFA, ESEM suffers from the issue of rotational indeterminacy, whereby all factor rotations that estimate the same number of factors will have exactly the same fit to the data (Asparouhov & Muthén, 2009; Marsh et al., 2014). The greater parsimony and flexibility of CFA makes it the clear favorite when ESEM and CFA measurement models have equivalent fit and parameter estimates (Asparouhov & Muthén, 2009; Marsh et al., 2014). Given the recency of this methodological development, the potential advantages of ESEM over traditional CFA approaches have yet to be fully explored in either the parenting or the psychopathology literature.

Factor Structure and Psychometric Invariance of ODD.

In Frick and colleagues' (1993) seminal meta-analysis of factor analytic studies containing both DSM-III and -III-R symptoms of ODD and CD, CD symptoms were grouped into three clusters while all ODD symptoms bunched together in a single cluster. This was interpreted as strong evidence of the unidimensionality of ODD, and was supported by behavior genetic studies demonstrating that unidimensional measures of ODD were etiologically distinct from other DBDs (Waldman, Rhee, Levy, & Hay, 2001; Dick, Viken, Kaprio, Pulkkinen, & Rose, 2005; Tuvblad, Zheng, Raine, & Baker, 2009) and general youth psychopathology (Lahey, Van Hulle, Singh, Waldman, & Rathouz, 2011; Cosgrove et al., 2011).

Recent investigations of ODD symptoms using traditional EFA and CFA models in the absence of other items have called this structure into question, and have found support for two (Rowe, Costello, Angold, Copeland, & Maughan, 2010; Lavigne, Gouze, Bryant, & Hopkins, 2014) and three factor models (Stringaris & Goodman, 2009; Burke, Hipwell, & Loeber, 2010a; Aebi et al., 2010; Burke, 2012; Krieger et al., 2013). This evidence is supported by behavior genetic findings supporting the presence of distinct etiologic influences in the oblique (correlated) two-factor model of ODD representing Irritability and Headstrong/Hurtful Behavior (Stringaris, Zavos, Leibenluft, Maughan, & Eley, 2012). These Irritability and Headstrong/Hurtful Behavior factors also appear to have distinctive phenotypic and etiologic relationships with depression and delinquency when measured prospectively in adolescence (Stringaris et al., 2012).

A major test of recently supported one-, two-, and three-factor models using CFA in five large independent samples (including the current sample) favored a modified bifactor structure, with an orthogonal general factor and two oblique specific factors representing Irritability and Oppositional Behavior (OB, Burke et al., 2014). These results were especially compelling due to the high convergence of results despite reliance on different assessment measures, different sample types, and different sample sex and age compositions (Burke et al., 2014). These results supported the hypothesis that a narrow irritability factor can be distinguished from the oppositional behaviors of ODD (Stringaris & Goodman, 2009), although average correlations between Irritability and OB were ~.80 (Burke et al., 2014). This high factor correlation casts some doubt on the discriminant validity of the Irritability and OB factors, especially since an orthogonal general factor was included (Burke et al., 2014). However, this poor discriminant validity may be a simple artifact of reliance on CFA.

While this modified bifactor model enjoyed general support, measurement and structural invariance across sample type, sex, and age could not be investigated (Burke et al., 2014). Prior studies have found invariance of some measurement parameters across sex and age among children (4-7) for a six-item oblique two-factor CFA model (Lavigne et al., 2014) and across mother-, father-, and teacher-report among Thai adolescents (7^{th}-12^{th} grade) and Spanish children (1^{st}-4^{th} grade) for an eight-item ESEM general factor model (Burns et al., 2013). Full measurement and structural invariance across sex for the same eight-item general factor model was supported in an earlier study of American (3-16) and Malaysian (5-12) youth using CFA (Burns, Walsh, Gomez, & Hafetz, 2006). No previous studies have examined full measurement and structural invariance across sex and age for the modified bifactor model identified by Burke and colleagues (2014). In addition, no prior studies have compared CFA and ESEM models. A re-examination of the factor structure of the EDRS ODD scale using ESEM may clarify the source of the high correlations between Irritability and OB (Burke et al., 2014). In addition, none of the recent factor analytic studies probing the potential sub-dimensions of ODD have examined their structural relations with parenting, or the invariance of these relations across sex and age.

Factor Structure and Psychometric Invariance of CD.

In Frick and colleagues' (1993) meta-analysis, CD could be decomposed into at least two major clusters often labeled Aggression and Rule-breaking. Rule-breaking could be further divided into property violations (e.g., vandalism) and status violations (e.g., truancy, Frick et al., 1993). Most follow-up work relying on factor analytic measurement models has focused on the two-factor model of CD (e.g., Tackett, Krueger, Sawyer, & Graetz, 2003; Tackett, Krueger, Iacono, & McGue, 2005; Burt, 2009; Burt, 2012; Burt & Klump, 2012; Burt, 2013; Burt, Donnellan, Iacono, & McGue, 2011), and this literature largely supports the distinction between Aggression and Rule-breaking, which appear to have distinct etiologic influences (Burt, 2009; Burt & Klump, 2012; Burt, 2013), developmental trajectories and rank-order stability (Tremblay, 2010), prevalence across sex (Moffitt, 2003), and external correlates (Burt, 2012; Lahey & Waldman, 2012). Although these potential differences in etiology, developmental trajectory, rank-order stability, prevalence across sex, and external correlates are intriguing, their strength is diminished due to a lack of studies testing the psychometric invariance of the measures used to operationalize CD and its sub-factors. For example, no prior studies examining differences in developmental trajectories and rank-order stability between Aggression and Rule-breaking have rigorously tested the assumption of age invariance. Similarly, limited information regarding sex invariance complicates the interpretation of sex differences in Aggression and Rule-breaking.

The Aggression and Rule-breaking factors of the CBCL, which include some items that are either not technically DSM symptoms of CD or are DSM symptoms of ODD, has shown evidence of weak, but not strong invariance across age using CFA in a U.S. twin sample spanning early childhood-late adolescence (Harden et al., 2015). In a sample of Mauritian 11 year-olds, the CBCL only satisfied configural invariance across sex and religion/ethnicity using CFA, unless the U.S.-developed factor structure was altered (including dropping the Rule-breaking factor), which achieved strong invariance (Yarnell et al., 2013). Strong invariance has been supported across sex for a one-factor model using CFA in an 11-symptom scale derived from a structured telephone interview of DSM-IV CD in a retrospective study of youth CD in two samples of adult twins (Meier, Slutske, Heath, & Martin, 2009). Without additional evidence of psychometric invariance of DSM-based measures of CD across sex and age, it is difficult to interpret the aforementioned findings, let alone the broader literature examining parenting influences on CD.

An additional threat to confidence in the distinction between Aggression and Rule-breaking lies in the moderate-to-high factor correlations that are typically found between Aggression and Rule-breaking across a variety of measures (avg. r = .55 (.28-.73), Burt, 2012), and in meta-analyses of commonly used measures, such as the CBCL (Achenbach 1991) and Youth Self-Report (YSR, Achenbach & Rescorla, 2001; Burt et al., 2015). As with ODD, these high factor correlations may simply be the result of heavy reliance on CFA, which dominates the factor analytic approach to studying CD. Previous

EFA analyses using the two-factor model have consistently demonstrated high item cross-loadings (e.g., Tackett et al., 2003). It is unknown whether use of an ESEM model would offer better discriminant validity.

Factor Structure and Psychometric Invariance of Parenting.

A consensus classification scheme for the highly multidimensional construct of parenting has yet to be achieved (O'Connor, 2002; McKee, Colletti, Rakow, Jones, & Forehand, 2008). Attempts at a broad classification system have generally focused on two orthogonal linear dimensions inspired by the work of Baumrind (1971), which are thought to underlie all specific parenting practices (Maccoby & Martin, 1983; Baumrind, 1991). The first dimension, responsiveness, reflects a parent's general level of sensitivity and responsiveness to child needs. The second dimension, demandingness, reflects a parent's general level of demand for child compliance and conformity (Maccoby & Martin, 1983; Baumrind, 1991). The few empirical tests of responsiveness support its conceptualization as a single linear dimension (e.g., Ten Haaf, Janssens, & Gerris, 1994). Demandingness, on the other hand, may be split into two sub-dimensions: psychological control and behavioral control (Barber et al., 1994). Psychological control represents a general approach to childrearing that intrudes on child psychosocial development, and includes tactics such as love withdrawal, maintaining child dependence, and guilt induction (Barber et al., 1994). Behavioral control refers to managing child behavior through setting and enforcing rules and oversight (Barber et al., 1994). The measurement strategy for these broad dimensions usually focuses on cross-situational attitudes and behaviors related to childrearing (Darling & Steinberg, 1993), and they are almost exclusively extracted from questionnaires.

These broad dimensions are contrasted with parenting practices, which are conceptualized as situation-specific attitudes and behaviors (Darling & Steinberg, 1993). Unlike the limited set of broad parenting dimensions, parenting practices are numerous and are captured using both self-report and observational measures (e.g., Reid, Patterson, & Snyder, 2002). The broad dimensions have been theorized to underlie all parenting practices and give rise to their affective valence, such that the same parenting practice can be delivered with a different “tone” based on a parent’s level of responsiveness, behavioral control, and psychological control (Darling & Steinberg, 1993; Barber, 1994). The impact of parenting behaviors, as well as the type of parenting behaviors emitted, may further depend on parent and child gender, and child age (e.g., Cote & Azar, 1997), although it is uncommon for studies of differential parental treatment within families to examine sex and age measurement invariance.

The current study relies on the parent self-report M-CRPR to measure the dimensions of Nurturance (responsiveness) and Restrictiveness (psychological control). In previous work, these two orthogonal factors have been consistently extracted using principal components analysis (PCA) and CFA in samples differing in socioeconomic status, geography, age, education, measurement format (Rickel and Biasatti, 1982) and culture (Deković et al., 1991). Restrictiveness has been shown to have positive relations with the use of spanking in early childhood (Holden, Coleman, & Schmidt, 1995), aggression among preschoolers (Stormont-Spurgin & Zentall, 1995), good school achievement among adolescents in high-risk neighborhoods and poor school achievement among adolescents in low-risk neighborhoods (Gonzales, Cauce, Friedman, & Mason, 1996), and antisocial behavior from early-to-mid-adolescence (Mason et al., 1996).

Nurturance has demonstrated a negative relation with child moodiness and learning problems among preschoolers (Atlas & Rickel, 1988), and a positive relation with school achievement in early adolescence (Gonzales et al., 1996). As with many measures of parenting, the two-factor Nurturance and Restrictiveness model of the M-CRPR has not been subjected to rigorous invariance testing and the M-CRPR has never been examined using ESEM.

The S/I scale focuses on specific parenting practices related to parental oversight and involvement (Loeber et al., 1998). The scale has been broken into two major constructs, Supervision and Involvement, using a rational-deductive approach (Loeber et al., 1998). Supervision items ask about parental tracking and knowledge of the child's whereabouts, activities, and associates when the child is away from direct parental oversight (Loeber et al., 1998). This set of parenting behaviors is usually labeled "monitoring" (Dishion & McMahon, 1998). Involvement items ask about the amount of discussions, planning, and shared activities between parent and child, and whether the parent and child enjoy their time together (Loeber et al., 1998).

The S/I scale has been used extensively in major longitudinal studies, such as the Pittsburgh Youth Study (Loeber et al., 1998), the Developmental Trends Study (Loeber, Green, Lahey, Frick, & McBurnett, 2000), and the Pittsburgh Girls Study (Keenan et al., 2010), which track the development of psychopathology and antisocial behavior among diverse populations of high risk or clinic-referred boys and girls. Despite its popularity and relevance to the etiology of antisocial behavior, the hypothesized internal structure of the S/I scale has never been examined using factor analysis, nor has it been the subject of any invariance testing. In previous work, Supervision has demonstrated significant negative relations with conduct problems from childhood-to-mid adolescence among boys (Pardini, Fite, & Burke, 2008) and joining a serious gang in early adolescence among boys (Lahey, Gordon, Loeber, Stouthamer-Loeber, & Farrington, 1999). Involvement has demonstrated significant negative relations with conduct problems from childhood-to-mid adolescence among boys (Pardini et al., 2008), precocious sexual behavior among early adolescent girls (Hipwell, Keenan, Loeber, & Battista, 2010), and with depression and CD severity from early-to-late adolescence among girls (Scott et al., 2013).

The Current Study

In the current analyses, data from the Georgia Twin Study, a large cross-sectional community sample of identical and fraternal twins aged 4-17 (Waldman et al., 1998), will be used to examine the factor structure of the M-CRPR, the S/I scale, and the EDRS ODD and CD scales. ESEM and CFA measurement models will be compared, and the best overall models will be used to test multigroup measurement and structural invariance across sex and age. If the minimum level of measurement invariance required for the valid use of latent variables cannot be rejected for each instrument, then the invariance of the linear effects of the broad parenting dimensions and specific parenting practices on youth ODD and CD will also be examined across sex and age. The incremental validity of the broad dimensions over the specific practices, and vice versa, will also be examined. Although non-linear and synergistic effects have been hypothesized (e.g., Darling & Steinberg, 1993; Caron, Weiss, Harris, & Catron, 2006), the current sample size is insufficient for offering any conclusions about such relations.

It is hypothesized that reanalyzing the EDRS ODD scale in the GTS sample using ESEM will yield lower factor correlations and better model fit due to the allowance of item cross-loadings. It is also hypothesized that the allowance of item cross-loadings in the ESEM model will eliminate the need for the orthogonal general factor, leaving an oblique Irritability and OB model. Given findings of sex, age, and/or informant measurement and structural invariance in previous work (e.g., Burns et al., 2006; Burns et al., 2013; Lavigne et al., 2014), it is hypothesized that full measurement and structural invariance across sex and age will not be rejected in the current sample.

It is predicted that the EDRS CD scale will match the two factor Rule-breaking and Aggression model that has been observed in previous investigations using alternative measures (Burt, 2012; Lahey & Waldman, 2012). Based on previous EFA analyses (e.g., Tackett et al., 2003), it is predicted that an ESEM model will be preferred over a CFA model due to better model fit and lower factor correlations. Based on limited prior research using DSM-based items (e.g., Meier et al., 2009), it is hypothesized that at least strong invariance will be replicated across sex.

It is hypothesized that the orthogonal two-factor model of Nurturance and Restrictiveness in the M-CRPR will be replicated in the current sample, and that an ESEM model will be preferred due to better model fit. There are no prior studies available to base predictions about invariance analyses.

It is hypothesized that the presumed two-factor structure of Supervision and Involvement in the S/I scale will match the internal structure derived from factor analysis, and that an oblique ESEM model will be preferred over an oblique CFA model due to the conceptual overlap of the items. No predictions can be made regarding sex and age invariance testing based on prior work, aside from likely age differences in factor means. Prior longitudinal research suggests that adolescents spend increasingly less time with their families as they move from age 10-18, but that time spent in discussion with parents remains stable (Larson, Richards, Moneta, Holmbeck, & Duckett, 1996).

Given that behavioral and psychological control may be distinct aspects of demandingness (Barber et al., 1994), it is hypothesized that correlations between Restrictiveness and Supervision will be low. Similarly, given the orthogonal nature of responsiveness and demandingness (Maccoby & Martin, 1983; Baumrind, 1991), it is hypothesized that correlations between Nurturance and Supervision will be low. Given the importance of parent-child relationship quality in the Involvement items (Loeber et al., 1998), it is assumed that Involvement will have moderate-to-high positive correlations with Nurturance and low magnitude associations with Restrictiveness.

Given prior positive relations between Restrictiveness and youth antisocial behavior (e.g., Gonzales et al., 1996; Mason et al., 1996; Stormont-Spurgin & Zentall, 1995), it is predicted that Restrictiveness will have positive relations with at least some aspects of ODD and CD. Given prior evidence suggesting positive relations of Nurturance (e.g., Atlas & Rickel, 1988; Gonzales et al., 1996) Supervision, and Involvement (e.g., Lahey et al., 1999; Pardini et al., 2008; Hipwell et al., 2010; Scott et al., 2013) with youth adjustment, it is assumed that these variables will have negative relationships with at least some aspects of ODD and CD.

Based on the theoretical predictions of Darling & Steinberg (1993), it is hypothesized that Nurturance and Restrictiveness, which are more distal variables, will have smaller linear relations than Supervision and Involvement, which are more proximal variables, for all child behavior constructs. Furthermore, it is hypothesized that Supervision and Involvement will still have significant linear relations independent of the relations of the more diffuse and distal Nurturance and Restrictiveness. All relations are assumed to be small-to-moderate, as is generally the case for parenting (e.g., Hoeve et al., 2009; Kaminski et al., 2008). It is unknown whether these parenting-DBD relations will be invariant across sex and age.

Method

Sample

The sample consisted of 846 twin pairs from the GTS, a non-referred population sample of twins born in Georgia between 1980 and 1991 (mean age = 10.6, *SD* = 3.2 years, age range = 4-17 years), with 49% males, 82% European Americans, 11% African Americans, 1% Hispanic Americans, and 6% mixed/other ethnicity. The sample contained 392 (46%) monozygotic (MZ) and 454 (54%) dizygotic (DZ) twin pairs. Mean household income was $53,000 (*SD* = $28,500). The sample was recruited according to the following procedures. In 1992 and 1993, 5,620 parents of child and adolescent twins born in Georgia were identified via state birth records and invited to join the Georgia Twin Registry by mail. Of the families contacted, 1,567 chose to join the Georgia Twin Registry, and 846 families (typically mothers) provided ratings on offspring psychopathology and parenting.

Twin zygosity was estimated using an 8-item questionnaire, which asked parents to rate their twins' level of physical similarity (e.g., "Are your twins as alike as two peas in a pod?") using a dichotomous rating scale (Bonnelykke, Hauge, Holm, Kristofferson,

& Gurtler, 1989). Item responses were averaged for each twin pair. Twins with a score of 0.5 or above were labeled MZ and twins with a score below 0.5 were labeled DZ. In previous research, the estimated internal consistency of this zygosity questionnaire in the GTS was good (α = .86) (Dong, Wu, & Waldman, 2014). Relative to direct measures of DNA similarity, this method of zygosity estimation has demonstrated at least 90% accuracy in distinguishing MZ and DZ twin pairs (Jackson, Sneider, Davis, & Treiber, 2001; Spitz et al., 1996).

Measures

Modified Child-Rearing Practices Report.

A 46-item parent-report version of the Modified Child-Rearing Practices Report (M-CRPR) (Rickel & Biasatti, 1982) was used to measure general parenting attitudes and behaviors related to children and child-rearing (Holden & Edwards, 1989). The items of this version of the M-CRPR are on a 5 category ordinal scale (0 = *not all* descriptive of how I raise my child; 4 = describes how I raise my child *very well*). The original CRPR (Block, 1965) was a 91-item Q-sort task estimated to have between 28-33 highly specific factors with moderate-to-low reliabilities (Block, 1973). Rickel and Biasatti (1982) reduced this Q-sort to a 40-item parent self-report questionnaire with 6 point ordinal items using principal components analysis (PCA) with Varimax rotation. Two orthogonal factors representing parental Nurturance (mean α = 0.80) and Restrictiveness (mean α = 0.76) were identified and demonstrated stability across samples differing in socioeconomic status, geography, age, education, measurement format (Rickel and Biasatti, 1982) and culture (Deković et al., 1991).

Supervision/Involvement Scale.

A 35-item parent report version of the S/I scale (Loeber et al., 1998) was used to index specific parenting practices. The original 43-item version of the scale was developed in three large samples of male children (1^{st}, 4^{th}, and 7^{th} graders) and relied on a combination of parent and child report. The Supervision and Involvement scales were comprised of two and four subscales, respectively. These subscales were extracted by examining item content, item inter-correlations, and subscale Cronbach's alphas (1951). Subscales for Supervision included *Poor Supervision* (mean $\alpha = 0.67$) and *No Set Time Home* (mean $\alpha = 0.59$). Subscales for Involvement included *Low Family Talk* (mean $\alpha = 0.76$), *Low Family Activities* (mean $\alpha = 0.73$), *Boy Not Involved* (mean $\alpha = 0.65$), and *Don't Enjoy Boy* (one item).

Both the original 43-item version of the S/I scale, and the 35-item version used in the GTS contain a mixture of interval, ordinal, and dichotomous items. The majority of items on the GTS version of the S/I scale are 5 category ordinal items (0 = *not all* descriptive of how I raise my child; 4 = describes how I raise my child *very well*). Two interval items asking about specific curfew times on weeknights and weekends were converted to 7 category ordinal items and reverse-scored (e.g., 6 = weekend curfew < 7pm, 0 = weekend curfew ≥ 12am; 6 = school night curfew < 6pm, 0 = school night curfew ≥11pm). Item 24 ("your child prefers to be with her/his friends rather than with the family") was also reverse-scored. Four interval items asking about the amount of waking hours spent each day with the child and the portion of that time spent in shared activities with the child were dropped from the analysis. These items clustered separately from all other items in initial EFAs, likely due to differences in level of measurement rather than substantive differences in item content.

Emory Diagnostic Rating Scale.

The EDRS was developed specifically for the GTS, to provide both continuous symptom ratings and diagnoses of DSM-IV childhood psychopathology, including ODD, CD, attention-deficit/hyperactivity disorder, major depression/dysthymia, generalized anxiety disorder, social phobia, simple phobia, separation anxiety disorder, panic disorder, agoraphobia, obsessive-compulsive disorder, tics and Tourette's disorder, and post–traumatic stress disorder (Waldman et al., 1998). Symptoms are measured on a 5 category ordinal scale asking parents how well each symptom has characterized their child over the past year (0 = *not at all*, 4 = *very well*), or in the case of 13 out of 15 CD symptoms, the frequency of symptoms over the past year (0 = *not at all*, 4 = *more than 3 times*). The ODD scale measures all 8 DSM-IV symptoms (e.g., "loses temper," "argues with adults") using 8 items, and the CD scale measures all 15 DSM-IV symptoms using 19 items (e.g., "has used a weapon that could cause serious harm," "skipped school or work"). The EDRS ODD and CD scales have demonstrated good reliability as single manifest scales (ODD, $\alpha = 0.91$; CD, $\alpha = 0.82$) in previous research using the GTS sample (Singh & Waldman, 2010; Ficks, Lahey, & Waldman, 2013).

Analysis

Estimation Method.

Mplus version 7.31 (Muthén & Muthén, 2012) was used for all analyses. The robust weighted least squares estimator (WLSMV) for ordinal item data was selected (Muthén & Muthén, 2012). This estimator relies on a matrix of polychoric correlations between all items (Flora & Curran, 2004). These polychoric correlations are based on the assumption that a continuous latent distribution underlies the observed responses to each ordinal item, and that the linear relationship between any two of these continuous latent response variables is bivariate normal (Flora & Curran, 2004). Previous CFA simulation research has shown that the WLSMV estimator provides accurate χ^2 test statistics, parameter estimates, and parameter standard errors when the continuous latent response variables underlying each ordinal item have moderate violations of normality (Flora & Curran, 2004). This simulation finding applies even in small samples with many items (Flora & Curran, 2004). WLSMV handles missing data using all the information available in the full sample when pairwise deletion is applied (Asparouhov & Muthén, 2010).

The *cluster* and *type = complex* options were selected to adjust model standard errors and χ^2 test statistics for the non-independence of observations within the sample (i.e. twins nested within the same families rated by the same parent) (Muthén & Muthén, 2012). The *difftest* option was used to conduct χ^2 difference tests adjusted for ordinal items and the WLSMV estimator. The *theta parameterization* was used in all analyses, so that item residual variances could be freely estimated in tests of multigroup strict invariance. The *grouping* option was selected to divide the sample into groups for tests of multigroup invariance. The *subpopulations* option was used to achieve correct standard errors when sex or age sub-samples were factor analyzed to assess the configural invariance of each factor structure identified in the full sample.

ESEM models were estimated with all rotated factor loadings freely estimated, and with the minimum constraints on the unrotated factor solution required for model identification (Asparouhov & Muthén, 2009). As in previous ESEM analyses, geomin rotation with $\varepsilon = .5$ was selected due to its excellent performance when: a) little is known about a scale's true loading structure, b) significant cross-loadings are present, c) model complexity is simple to moderate, and d) it cannot be assumed that one item loading per factor = 0 (Asparouhov & Muthén, 2009; Marsh, Nagengast, & Morin, 2013; Marsh et al., 2014). Oblique geomin rotation was used to allow maximum flexibility in the exploration of factor structure.

Goodness-of-fit.

Model fit indices used included the χ^2, the χ^2 difference test (Loehlin, 2004), the root mean square error of approximation (RMSEA, Steiger, 1990), the root mean square error of approximation lower bound index (RMSEA.LB, Preacher, Zhang, Kim, & Mels, 2013), the comparative fit index (CFI, Bentler, 1990), and the Tucker-Lewis index (TLI, Tucker & Lewis, 1973).

The χ^2 and its associated degrees of freedom (df) provide an estimate of absolute model fit yielding a p-value, with p-values greater than .05 indicating no significant differences between the observed and expected variance-covariance matrices (Loehlin, 2004). When two hierarchically nested models are compared (i.e. the parameters of one model can be constrained to produce the simpler alternative model), a χ^2 difference test ($\Delta\chi^2$) can be performed using the χ^2 and df for each model, yielding a p-value that indicates whether the two models are significantly different in fit ($p < .05$; Loehlin, 2004). Unfortunately, the χ^2 and $\Delta\chi^2$ test statistics are sensitive to sample size, such that minor discrepancies between the hypothesized model and the observed data, or between two hierarchically nested models, are more likely to generate a p-value below .05 as sample size increases (Loehlin, 2004). Therefore, while χ^2 and $\Delta\chi^2$ statistics are reported in the tables below, the supplemental fit indices including the RMSEA, RMSEA.LB, CFI, and TLI were relied upon for model evaluation as these are less sensitive to Karegeannes samples (Hu & Bentler, 1999; Preacher et al., 2013).

Following established guidelines, RMSEA ≤ .08 and ≤ .06 imply adequate and excellent model fit, respectively (Hu & Bentler, 1999) and CFI and TLI ≥ .90 and ≥ .95 imply adequate and excellent model fit, respectively (Hu & Bentler, 1999). When comparing hierarchically nested models, established guidelines suggest that increases of RMSEA ≤ .015, and decreases in CFI and TLI ≤ .01, indicate no meaningful decrement in model fit following parameter constraint (Chen, 2007; Cheung & Rensvold, 2002). Changes in RMSEA, CFI, and TLI are represented as ΔRMSEA, ΔCFI, and ΔTLI.

When the factor structure of a questionnaire is unknown, ESEM can be used in an exploratory manner, and the RMSEA.LB index may be advantageous in guiding the selection of the optimal number of factors (Preacher et al., 2013). The RMSEA.LB index selects the model with the fewest number of factors required to bring the lower bound of the 90% confidence interval for the RMSEA below .05 (Preacher et al., 2013). This increases the RMSEA's preference for model parsimony. A recent EFA simulation demonstrated that the RMSEA.LB index performed better than the RMSEA, Akaike Information Criterion (AIC, Akaike, 1987), and Bayesian Information Criterion (BIC, Schwartz, 1978) in recovering the "true" factor structure, regardless of sample size (Preacher et al., 2013). As a result, the RMSEA.LB index was given special consideration in exploratory ESEM, although all other fit indices, model interpretability, factor discriminant validity, consistency across sex and age sub-samples, scale score reliability, and prior research were also considered.

Factor Analysis and Measurement/Structural Invariance.

Factor analysis. Factor solutions generated using CFA were compared to factor solutions generated using ESEM. When ESEM and CFA models were approximately equivalent, the more parsimonious CFA model was always preferred (Marsh et al., 2014). CFA and ESEM factor analyses focused primarily on first order oblique or orthogonal models, although CFA bifactor models (Reise, 2012) were included when indicated by prior research (Burke et al., 2014). Bifactor models assume that all items load onto an orthogonal general factor in addition to their loadings on orthogonal specific factors, and this general factor is assumed to account for all shared variance among the specific factors (Reise, 2012), although oblique specific factors were allowed based on prior work (Burke et al., 2014).

Scale score reliability was calculated using McDonald's (1970) omega ($\omega = (\Sigma\square|\lambda i|)^2 / ([\Sigma\square|\lambda i|]^2 + \square\delta ii)$) where λi represents standardized factor loadings and δii represents standardized item residual variance (McDonald, 1978). Unlike Cronbach's coefficient alpha (α, 1951), which relies on the assumption that all items in a scale are equally good indicators of its underlying construct (tau-equivalence) (Sijtsma, 2009), McDonald's ω (1970) relies on the absolute magnitudes of factor loadings and item residual variances, allowing it to generate more accurate lower bound reliability estimates when tau-equivalence is violated (McDonald, 1978). McDonald's ω (1970) is well suited for multidimensional measurement models (McDonald, 1978). Given the greater familiarity of Cronbach's α (1951), α coefficients were also reported. In the early stages of test development in basic social science research, reliability estimates $\geq .70$ are traditionally considered acceptable and estimates $\geq .80$ are traditionally considered good

(Lance, Butts, & Michels, 2006). In applied settings, estimates ≥ .90 have been recommended (Lance et al., 2006).

Multigroup measurement/structural invariance. Selecting factor solutions that were consistent across sex and age during factor analysis ensured Configural invariance. For sex invariance analyses, the sample was split into males and females. For age invariance analyses, the sample was split at the median age (10.38). An age 10.38 split roughly corresponds to an age 10 split, which is the earliest accepted cut point between childhood and adolescence (Smetana, Campione-Barr, & Metzger, 2006) The large sample size of the GTS buffered against the loss of statistical power due to dichotomizing age.

Baseline configural invariance models were established according to the process described by Guay and colleagues (2014). Next, weak, strong, strict, factor variance-covariance, and factor mean invariance was added incrementally to the model, respectively (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011; Guay et al., 2014). Lack of invariance was suggested if a meaningful decrement in model fit by ΔRMSEA, ΔCFI, and ΔTLI was observed when a new level of invariance was added to the model (Chen, 2007; Cheung & Rensvold, 2002).

If invariance was rejected at any step, modification indices (M.I.) were requested in the *output* section of the Mplus syntax using the *modindices (all)* command to identify which parameters needed to be freely estimated across groups (Muthén & Muthén, 2012; Marsh et al., 2013). Model M.I.s indicate the expected change in model χ^2 following a change to an identified model parameter, with a χ^2 change of 3.84 (M.I. = 3.84) indicating statistical significance (Muthén & Muthén, 2012). An M.I. of 10 was used as the minimum necessary to consider a post-hoc change to a model parameter, as suggested by Muthén and Muthén (2012). If instances of parameter non-invariance were minimal and potentially meaningful, invariance testing proceeded for the remaining model parameters with the identified instance(s) of parameter non-invariance included in the model (Byrne, Shavelson, & Muthén, 1989). Partial invariance could be implemented for item thresholds, item residual variances, factor variances, and factor means in both ESEM and CFA (Asparouhov & Muthén, 2009, Marsh et al., 2014). Partial invariance of item factor loadings and factor co-variances could only be implemented in CFA due to the use of factor rotation in ESEM (Asparouhov & Muthén, 2009, Marsh et al., 2014).

MIMIC measurement/structural invariance. Due to little variation in item responses for many CD items, a multiple indicator multiple causes (MIMIC) model (Jöreskog & Goldberger, 1975; Muthén, 1989; Marsh, Tracey, & Craven, 2006; Marsh et al., 2013) was selected to examine at least item threshold and factor mean invariance across sex and age. In MIMIC modeling, item threshold and factor mean invariance is examined by regressing these parameters on selected grouping variables simultaneously using the full sample (Jöreskog & Goldberger, 1975; Muthén, 1989; Marsh et al., 2006; Marsh et al., 2013). In theory, this approach has the flexibility to allow synergistic and higher order sex and age effects, but only sex and age linear paths were examined to preserve model identification (Jöreskog & Goldberger, 1975; Muthén, 1989; Marsh et al., 2006; Marsh et al., 2013). Key assumptions of this approach are the presence of configural and weak invariance, and uniform differences between thresholds across the entire continuous latent response distribution (Marsh et al., 2006; Marsh et al., 2013). Despite limitations in testable measurement parameters, MIMIC models enjoy greater statistical power than multigroup models for continuously measured grouping variables like age (Muthén, 1989; Marsh et al., 2006; Marsh et al., 2013).

In order to preserve model identification, the null model was tested first by setting all sex and age linear paths on items and factors to zero. Model M.I.s were requested and inspected to see if freeing age and/or sex linear paths on specific items and/or factors could create significant improvements in model fit (Muthén, 1989; Marsh et al., 2006; Marsh et al., 2013). A new model was then tested, freely estimating the age and/or sex linear paths with the most substantial M.I.s (Muthén, 1989; Marsh et al., 2006; Marsh et al., 2013). This less invariant model was compared to the null model by ΔRMSEA, ΔCFI, and ΔTLI, and significant improvements in model fit were interpreted as evidence of partial non-invariance (Marsh et al., 2006; Marsh et al., 2013). This iterative process guided by M.I.s continued until no further significant improvements in model fit could be achieved (Marsh et al., 2006; Marsh et al., 2013). This iterative process should be interpreted with caution, as it may unduly capitalize on chance.

Structural Equation Modeling.

ODD analyses were conducted using the sex and age multigroup invariance models developed earlier in the analysis. CD analyses were conducted using full sample MIMIC versions of the sex and age multigroup invariance models developed earlier in the analysis for all parenting scales.

The invariance of unstandardized path coefficients (b) across sex and age for ODD was tested by comparing a multigroup model in which path coefficients were freely estimated across groups to a multigroup model in which path coefficients were constrained to equality across groups (Millsap, 2007). ΔRMSEA, ΔCFI, and ΔTLI below their cut-off values were considered evidence of path coefficient invariance. The incremental validity of the specific parenting practices over the broad parenting dimensions were tested by examining changes in R^2 for the ODD factors when the specific practices were added to the broad dimensions model, and vice versa.

The linear relations of the parenting variables and CD were tested following a similar sequence, without the tests of path coefficient invariance across sex and age. The CD MIMIC model residualized on linear sex and age terms served as the baseline.

In order to allow tests of partial path coefficient invariance, ESEM measurement models were transformed into a CFA format using the ESEM-within-CFA (ES-W-C) method described by Marsh and colleagues (2013) prior to being added to the SEM analysis. This produces a CFA model with the same degrees of freedom, χ^2 value (within rounding error), fit statistics, and parameter estimates as found in the ESEM model (Marsh et al., 2013). When needed, additional key parameters such as factor correlations were fixed to their values in the ESEM model to ensure the consistency of the measurement models across SEM analyses.

Due to a lack of guidelines for the qualitative description of β as effect sizes, β were labeled “small” if $.10 \leq \beta < .30$, “medium” if $.30 \leq \beta < .50$, and “large” if $\beta \geq .50$, in keeping with Cohen’s (1992) definitions for bivariate *r*.

Results

Factor Analysis and Measurement/Structural Invariance.

Modified Child-Rearing Practices Report.

***Factor analysis*.** Exploratory ESEM with up to eight factors identified an oblique two-factor Nurturance and Restrictiveness model as best based on fit, interpretability, consistency across the full sample and sub-samples, scale score reliability, and prior research (Rickel & Biasatti, 1982; Deković et al., 1991) (Tables 1 and 2). Factor loadings and factor correlations were consistent across the full sample and sub-samples in pattern, magnitude, and direction (Table 2).

In approximate terms, model fit was excellent by RMSEA (ranged from .037-.040), and inadequate by CFI (ranged from .806-.849) and TLI (ranged from .787-.834) (Table 1). In the full sample and sub-samples, ΔRMSEA did not indicate a meaningful decrement in model fit with the progressive removal of factors, until the model was reduced from two to one factor. By RMSEA.LB, the two-factor model fit best in the full (RMSEA 90% CI [**.037**, .041]), male (RMSEA 90% CI [**.036**, .042]), female (RMSEA 90% CI [**.037**, .043]), child (RMSEA 90% CI [**.032**, .038]), and adolescent sub-samples (RMSEA 90% CI [**.034**, .040]). In contrast, χ^2, CFI, TLI, and RMSEA favored the most complicated model tested in the full sample and sub-samples. This eight-factor model had poor interpretability. This is consistent with the tendency of χ^2, CFI, TLI, and RMSEA to over-factor, particularly when there are many items and a large sample (Preacher et al., 2013; Guay et al., 2014). ΔCFI, and ΔTLI also preferred the eight-factor model in the full sample and sub-samples, despite its poor interpretability.

Coefficients of congruence (Watkins, 2002) with the 18-item Nurturance and 20-item Restrictiveness standardized factor loadings from Rickel & Biasatti (1982) in the current full sample and sub-samples were excellent (Nurturance: r_c = .97-.98) to good (Restrictiveness: r_c = .95-.97, MacCallum, Widaman, Zhang, & Hong, 1999) (Table 2).

This high congruence was observed even though a different method of factor extraction (PCA) and rotation (Varimax), and a slightly different questionnaire were used in Rickel & Biasatti (1982). The oblique geomin rotation in the current analysis was also able to replicate the orthogonal nature of Rickel & Biasatti's (1982) Nurturance and Restrictiveness factors (average $r = .02$, ns at $p <.05$). McDonald's (1970) ω scale score reliability estimates for the two-factor model were good in both the full sample and sub-samples (Nurturance: $\omega = .84\text{-}.91$; Restrictiveness: $\omega = .84\text{-}.85$) (Table 2). Cronbach's α (1951) scale score reliability estimates were also good (Nurturance: $\alpha = .80\text{-}.85$; Restrictiveness: $\alpha = .79\text{-}.81$) (Table 2).

The two-factor oblique ESEM model was compared to a two-factor oblique CFA model with an identical pattern of target loadings and all non-target loadings fixed to zero (Tables 2 and 3). Factor loadings and factor correlations for the full sample ESEM and CFA models were very similar in magnitude and direction (Table 2). The CFA model also had excellent discriminant validity and good scale score reliability (Nurturance: full sample $\omega = .91$, $\alpha = .82$; Restrictiveness: full sample $\omega = .85$; $\alpha = .80$) (Table 2). However, the CFA model fit was relatively poor across the full sample and sub-samples: ΔCFI (ranged from -.06 to -.114) and ΔTLI (ranged from -.055 to -.112) (Table 3). This decrement in fit likely reflects the item cross-loadings that were fixed to zero in the CFA model (full sample: 27 items; 5 items with cross-loadings ≥ .2). In order to reduce model misspecification, the two-factor ESEM model was favored over the two-factor CFA model.

Multigroup measurement/structural invariance across sex. As can be seen in Table 4, full measurement and structural invariance could not be rejected. The final sex invariant factor loadings, factor correlations, and reliability estimates are listed in Table 5. Scale score reliability of the sex invariant model was good (Nurturance: ω = .91, α = .83; Restrictiveness: ω = .85, α = .80) (Table 5).

Multigroup measurement/structural invariance across age. With the exception of item 13's thresholds and residual variance (“I do not allow my child to say bad things about her/his teachers”), full measurement and structural invariance across offspring age could not be rejected (Table 4). Inspection of the standardized threshold estimates for item 13 when a partial strong invariance model was generated revealed that the thresholds for parents to endorse “I do not allow my child to say bad things about her/his teachers” were higher for parents of adolescents, suggesting that parents more readily reported restricting offspring expression of negative comments towards teachers if their offspring were children. The final age invariant factor loadings, factor correlations, and scale score reliability estimates are listed in Table 5. Scale score reliability of the age invariant model was good (Nurturance: ω = .91, α = .83; Restrictiveness: ω = .85, α = .80) (Table 5).

Supervision/Involvement Scale.

Factor analysis. Due to little variation in item responses, two dichotomous items - 12a (“Is s/he usually supervised? [in the evening]”) and 13a (“Is s/he usually supervised? [on weekends]”) – had to be dropped, leaving a 29-item scale. Exploratory ESEM of models with up to eight factors identified an oblique three-factor Monitoring, Direct Supervision, and Involvement model as best by fit, interpretability, consistency across the full sample and sub-samples, scale score reliability, and prior research (Tables 6 and 7). Factor loadings and correlations were consistent in pattern, magnitude, and direction across the full sample and sub-samples (Table 7). The key distinction between Monitoring, which corresponded closely to Loeber and colleagues' (1998) original Supervision scale, and Direct Supervision, appeared to be that Monitoring items focused on the process of tracking child behaviors while the child was outside of parental view while Supervision items focused on ensuring direct oversight of the child (Table 7). The items of Involvement corresponded almost exactly with the items of the original Involvement scale (Loeber et al., 1998) (Table 7).

In approximate terms, model fit was excellent by RMSEA (ranged from .044-.057), adequate by CFI (ranged from .908-.936) and TLI (ranged from .906-.920 in the male, female, and adolescent sub-samples), and near adequate by TLI (ranged from .884-894 in the full sample and child sub-sample) (Table 6). In the full sample and sub-samples, ΔRMSEA did not indicate a meaningful decrement in fit with the progressive removal of factors, until the model was reduced from three to two factors. By RMSEA.LB, the three-factor model fit best in the male (RMSEA 90% CI [**.046**, .055]), female (RMSEA 90% CI [**.049**, .058]), and child sub-samples (RMSEA 90% CI [**.039**, .049]), and nearly fit best in the full sample (RMSEA 90% CI [**.051**, .057]) and the adolescent sub-sample (RMSEA 90% CI [**.053**, .062]). Although the RMSEA.LBs in the full sample and the adolescent sub-sample were just beyond the cut-off (lower bound < .05), technically favoring the four-factor model (full: RMSEA 90% CI [**.045**, .051]; adolescent: RMSEA 90% CI [**.043**, .054]), this fourth factor worsened model interpretability in both cases. Consistent with their tendency to over-factor (Preacher et al., 2013; Guay et al., 2014), the χ^2, CFI, TLI, and RMSEA favored the most complicated model tested in the full sample and sub-samples. This eight-factor model had poor interpretability and factors with few item target loadings. ΔCFI and ΔTLI fit indices were inconsistent. McDonald's (1970) ω scale score reliability estimates for the three-factor model were good in both the full sample and sub-samples (Monitoring: ω ranged from .87-.89; Direct Supervision: ω ranged from .76-.85; Involvement: ω ranged from .91-.93) (Table 7). Cronbach's α (1951) scale score reliability estimates were generally acceptable (Monitoring: α ranged from .72-.78; Direct Supervision: α ranged from .55-.73; Involvement: α ranged from .84-.89) (Table 7).

This oblique three-factor ESEM model was compared to an oblique three-factor CFA model, in which all non-target factor loadings were fixed to zero (Tables 7 and 8). The CFA model fit worse in the full sample and sub-samples by ΔCFI (ranged from -.053 to -.075) and ΔTLI (ranged from -.042 to -.069) (Table 8). It also fit worse by ΔRMSEA in the male (+.015), female (+.017), and adolescent sub-samples (+.015), and was close to the cut-off in the full sample (+.013). In addition, the factor correlations for the ESEM model were lower than their counterparts in the CFA model (full sample: ESEM, .16, .20, .30; CFA, .36, .47, .46, respectively), suggesting better discriminant validity (Table 7). The ESEM model was selected as best overall.

Multigroup measurement/structural invariance across sex. As can be seen in Table 9, full measurement and structural invariance could not be rejected. The final sex invariant factor loadings, factor correlations, and reliability estimates are listed in Table 10. Scale score reliability of the sex invariant model was good (Monitoring: ω = .89, α = .74; Direct Supervision: ω = .84, α = .72; Involvement: ω = .93, α = .87).

Multigroup measurement/structural invariance across age. As can be seen in Table 9, no meaningful decrements or improvements in model fit by ΔCFI, ΔTLI, and

ΔRMSEA were observed in the progression from configural to factor variance-covariance invariance. In contrast, invariance of factor means for Monitoring, Direct Supervision, and Involvement across offspring age was rejected due to meaningful decrements in model fit by ΔCFI (-.038) and ΔTLI (-.037). Inspection of the standardized estimates of factor means at the factor variance-covariance step revealed that, on average, parents of adolescents reported more Monitoring (*d* = 52), less Direct Supervision (*d* = -1.42), and less Involvement (*d* = -.31) than parents of children. The final age invariant factor loadings, factor correlations, and reliability estimates are listed in Table 10. Scale score reliability estimates for this age invariant model were good (Monitoring: ω = .90, α = .74; Direct Supervision: ω = .80, α = .72; Involvement ω = .93, α = .87).

Emory Diagnostic Rating Scale – ODD.

Factor analysis. Oblique two- and three-factor ESEM models were compared. Based on interpretability, consistency across the full sample and sub-samples, scale score reliability, and prior research, the oblique two-factor model was selected as the best overall ESEM model.

No meaningful differences in model fit between the two- and three-factor ESEM models were observed based on ΔCFI, ΔTLI, and ΔRMSEA, except for an improvement of model fit with the elimination of the third factor in the female sub-sample by ΔRMSEA (-.018) (Table 11). In approximate terms, across the full sample and sub-samples, both the three- and two-factor models had excellent fit by CFI (three-factor: .992-.996; two-factor: .983-.993) and TLI (three-factor: .966-.984; two-factor: .964-.985), and adequate-to-inadequate fit by RMSEA (three-factor: .081-.124; two-factor: .079-.128).

Both the three- and two-factor models suffered some instability in the pattern of target and non-target factor loadings across the full sample and sub-samples, but the pattern in the two-factor model was more consistent and interpretable. This interpretability can be seen in the full sample results in Table 12. Three items preferentially loaded on the first factor including “loses temper,” “argues with adults,” and “disobeys adults.” This factor corresponded to the ODD Behavior factor identified by Burke and colleagues using EFA (2010a) and replicated by Lavigne and colleagues (2014) using ICM-CFA. It was labeled ODD Behavior (ODDB). Five items preferentially loaded on the second factor, which appeared to collapse the ODD Negative Affect (“touchy,” “angry,” and “spiteful/vindictive”) and ODD Antagonistic (“annoys others,” and “blames others”) factors identified by Burke and colleagues (2010a) using EFA into a single factor. This factor was labeled ODD Negative/Antagonistic (ODD N/A).

Scale score reliability estimates for this two-factor model were good (full sample: ODDB $\omega = .86$, $\alpha = .84$; ODD N/A $\omega = .91$, $\alpha = .90$). Despite the allowance of item cross-loadings, factor correlations were moderately high (e.g., full sample: $r = .72$).

The two-factor ESEM model was compared to the CFA modified bifactor model previously identified in the GTS sample by Burke and colleagues (2014) (Tables 11 and 12). The CFA modified bifactor model enjoyed better model fit by ΔRMSEA in the full sample (+.027), and was right at the cut-off for significantly better fit by ΔRMESA in the male (+.015) and female sub-samples (+.015) (Table 11). The CFA modified bifactor model was also at the cut-off point for better fit by ΔTLI (-.01) in the full sample. Similar to the two-factor ESEM model, the CFA modified bifactor model had excellent approximate fit by CFI (ranged from .989-.995) and TLI (ranged from .972-.988), and adequate-to-inadequate approximate fit by RMSEA (ranged from .071-.113). Although differences in model fit favoring the CFA modified bifactor model were not stark, the pattern of factor loadings was more easily interpreted in the CFA model and more consistent across the full sample and sub-samples (Table 12). In addition, the CFA model has already been cross-validated in four large, independent, and diverse samples employing different measures (Burke et al., 2014). The ESEM model also did not offer much improvement in factor discriminant validity (Table 12). Therefore, the CFA modified bifactor model was selected as the best overall model. Scale score reliability estimates for the model were generally acceptable (General: ω = .79-.96, α = .92-.93; OB: ω = .56-.94, α = .88-.89; Irritability: ω = .75-.92, α = .82-.87).

Multigroup measurement/structural invariance across sex.

When the baseline configural invariance model was initially established, residual variance for item 2 (“argues with adults”) was estimated as negative in the female group (δ = -16, p =.81), a so-called “Heywood case” (Chen, Bollen, Paxton, Curran, & Kirby, 2001). Although negative residual variance can indicate a poorly specified model, when it is non-significant, it can be attributed to simple sampling variability, and fixed to zero (Chen et al., 2001). Fixing this parameter did not meaningfully change model fit ($\Delta\chi^2$ = 5, p = .024; ΔCFI = .000; ΔTLI = .000; ΔRMSEA = +.001). Therefore, item 2 residual variance was fixed to zero in the female group for all stages of invariance testing. This resulted in the non-invariance of item 2 residual variance across sex, since item 2 residual variance had to be fixed to one in the male group for identification of the configural model.

Aside from this single instance of non-invariance, full measurement and structural invariance across sex could not be rejected (Table 13). In addition, inspection of standardized residual variance estimates for item 2 in the final measurement model revealed little substantive differences across sex (male: residual variance = 4%, 95% CI [-9%, 17%]; female: residual variance fixed to 0). The final sex invariant factor loadings, factor correlations, and reliability estimates are listed in Table 14. Scale score reliability estimates for the sex invariant model were good (General: ω = .93-.94, α = .93; OB: ω = .88-.89, α = .88; Irritability: ω = .86, α = .85).

Multigroup measurement/structural invariance across age.

When the baseline configural invariance model was initially established, residual variance for item 6 (“is touchy or easily annoyed by others”) was estimated as negative in the adolescent offspring group (δ = -3, p =.55). Fixing this parameter did not meaningfully change model fit ($\Delta\chi^2$ = 3, p = .081; ΔCFI = .000, ΔTLI = .000, ΔRMSEA = -.001). Therefore, item 6 residual variance was fixed to zero in the adolescent offspring group for all stages of invariance testing.

During strong invariance testing, an additional source of negative, non-significant item residual variance was uncovered in the adolescent offspring group - item 2 (“argues with adults;” δ = -60, p =.99). Fixing this parameter to zero in the adolescent offspring group did not meaningfully impact model fit ($\Delta\chi^2$ = 1, p = .28, ΔCFI = .000, ΔTLI = .000, ΔRMSEA = .000). Therefore, item 2 residual variance was fixed to zero in the adolescent offspring group in all further tests of invariance.

With these local instances of non-invariance incorporated in the model, the hypothesis of measurement and structural invariance across age could not be rejected overall (Table 13). Inspection of standardized residual variance estimates for item 2 in the final measurement model revealed little substantive differences across age (child: residual variance = 6%, 95% CI [-6, 18]; adolescent: residual variance fixed to zero). Differences were larger in magnitude for item 6 (child: residual variance = 11%, 95% CI [1-22]; adolescent: residual variance fixed to zero), although the lower bound of the 95% confidence interval was barely above zero.

The final age invariant factor loadings, factor correlations, and reliability estimates are listed in Table 14. Scale score reliability estimates for the age invariant model were acceptable (General: $\omega = .96$, $\alpha = .93$; OB: $\omega = .71$-$.72$, $\alpha = .88$; Irritability: $\omega = .73$-$.77$, $\alpha = .85$).

Emory Diagnostic Rating Scale – CD.

Factor analysis. Oblique three- and two-factor ESEM models of CD symptoms were compared. Fourteen items representing 10 CD symptoms had sufficient variation for factor analysis, and 11 items representing 7 CD symptoms had sufficient variation to be used in further SEM analyses with the parenting factors. The 11-item version of the scale had to be used. Four of the items in the 11-item scale represented various aspects of the symptom “often lies” and two of the items represented varying aspects of the symptom “often initiates physical fights” (APA, 1994). It was decided a priori to allow correlated errors for items representing the same kinds of symptom. Based on model fit, interpretability, scale score reliability, and prior research, the two-factor model was selected as the best overall ESEM model.

The three-factor ESEM model fit better than the two-factor ESEM model by ΔRMSEA (+.018), but there was little difference based on ΔCFI or ΔTLI (Table 15).

Both models had excellent approximate fit by CFI (three-factor: 1.00; two-factor: .998), TLI (three-factor: 1.00; two-factor: .996), and RMSEA (three-factor: .006; two-factor: .024). The three-factor model also demonstrated excellent absolute fit by χ^2 (p = .42). The close fit of the three-factor model seemed to represent model over-fitting, as the pattern of factor loadings in this model were difficult to interpret. The more parsimonious two-factor model had much better interpretability, and largely conformed to previous Aggression and Rule-breaking models generated using EFA (e.g., Tackett et al., 2003) (Table 16). Scale score reliability estimates for the two-factor model were good (Rule-breaking ω = .84, α = .60; Aggression ω = .87, α = .74) (Table 16). The correlation between Rule-breaking and Aggression (r = .42) was generally below the range reported in previous CFA studies relying on parent-report (r = .56-.73, Burt, 2012; r = .49-.58, Burt et al., 2015) (Table 16).

A series of nested and non-nested CFA models were generated to find the optimal model to compare to the two-factor ESEM model. Given high cross-loadings in the ESEM model, a modified bifactor model, with a single general factor and oblique Aggression and Rule-breaking specific factors, and a conventional bifactor model, with a single general factor and orthogonal Aggression and Rule-breaking specific factors, were generated. Two-factor oblique and orthogonal Aggression and Rule-breaking models, and a general factor model were also generated. First order three-factor Aggression, Property Violations, and Status Violations models (Frick et al., 1993) were not identified, likely due to the loss of half the CD symptoms. Selection of items to represent Aggression and Rule-breaking were determined a priori, based on item face validity and prior research (e.g., Tackett et al., 2003; Tackett et al., 2005; Burt et al., 2011). The bifactor models had the best fit statistics, but poor interpretability (Table 15). For example, in the modified bifactor model, the correlation between Aggression and Rule-breaking was estimated to be greater than one. In the bifactor model, half of the loadings on the Aggression factor were negative, while the other half were positive. These issues suggested model misspecification. In contrast, the oblique Aggression and Rule-breaking model had good interpretability (Table 16). In addition, this oblique two-factor model had excellent approximate fit by RMSEA (.053), CFI (.987), and TLI (.979) and fit better than the orthogonal Aggression and Rule-breaking model by ΔCFI (-.226), ΔTLI (-.334), and ΔRMSEA (+.167) (Table 15). No significant differences in fit were observed between this oblique two-factor model and the general factor model, even by the overly sensitive $\Delta\chi^2$ ($p = .12$). This likely reflects the excessively high correlations between Aggression and Rule-breaking in the two-factor oblique model ($r = .94$) (Table 16). The interpretability of this general factor model was also excellent, with standardized factor loadings ranging from .55-.85 (Table 16). However, based on prior findings supporting the distinction between Aggression and Rule-breaking (e.g., Moffitt, 2003; Burt, 2009; Tremblay, 2010; Burt, 2012; Burt & Klump, 2012; Lahey & Waldman, 2012; Burt, 2013), the two-factor oblique model was preferred. Scale score reliability estimates for both models were good (Aggression: $\omega = .86$, $\alpha = .64$; Rule-breaking = .87, $\alpha = .79$; General factor: $\omega = .92$, $\alpha = .82$).

Regardless of whether the general factor or oblique two-factor CFA model was selected, neither compared favorably in terms of model fit and/or discriminant validity to the oblique two-factor ESEM model. This two-factor ESEM model was selected as best overall.

MIMIC measurement/structural invariance: sex and age. The null model had excellent fit by CFI (.982), TLI (.971), and RMSEA (.054) (Table 16). Modification indices (M.I.) suggested age effects (M.I. = 64) on the mean of the Aggression factor, sex effects (M.I. = 34) on the mean of the Rule-breaking factor, and age effects on the item thresholds for “skipped school or work” (M.I. = 39) and “stayed out late against parents’ wishes” (M.I. = 31). A threshold invariance model was generated next, in which the sex and age paths to all factor means were freely estimated and the sex and age paths to all item thresholds were fixed to zero. This threshold invariance model fit better than the null model by ΔCFI (+.014), ΔTLI (+.021), and ΔRMSEA (-.026). Model M.I.’s were re-examined, and only one item continued to appear non-invariant across offspring age (“stayed out late against parents’ wishes”; M.I. = 13). Inclusion of this single instance of threshold non-invariance did not change model fit by ΔCFI (.000), ΔTLI (.000), and ΔRMSEA (-.001). Therefore, the hypothesis of item threshold invariance could not be rejected, but meaningful differences in factor means based on sex and age were detected. The standardized path coefficients of sex and age on the factor means are listed in Table 17.

The mean of parent-reported offspring Aggression in the female group was .19 (95% CI [-.09, -.29]) standard deviations lower than the mean of parent-reported offspring Aggression in the male group. There were no significant differences between male and female offspring groups in Rule-breaking (β = -.02, 95% CI [-.14, .10]). A one standard deviation increase in age was estimated to correspond to a .33 (95% CI [-.23, -.43]) standard deviation decline in Aggression. In contrast, a one standard deviation increase in age was estimated to correspond to a .40 (95% CI [.28, .52]) standard deviation increase in Rule-breaking behavior.

Structural Equation Models.

Broad Parenting Dimensions Predicting ODD.

Nurturance and Restrictiveness: Sex invariance. The fit of the invariant model showed no change in model fit by ΔCFI (+.005) ΔTLI (+.005), and ΔRMSEA (-.001) (Table 18). In addition, this invariant model had excellent fit by CFI (.927), TLI (.929), and RMSEA (.028). Therefore, invariance of the linear paths of Nurturance and Restrictiveness on the General, Irritability, and OB factors of ODD across offspring sex could not be rejected. The standardized path coefficients are listed in Table 19. Overall, the fully sex invariant model with all broad parenting dimensions included was estimated to explain 2% of the variance in the General factor (95% CI [.04, 4]), 7% of the variance in the Irritability factor (95% CI [1, 13]), and 10% of the variance in the OB factor (95% CI [4, 16]) (Table 19).

Nurturance was estimated to have small negative relations with the General (β = -.12, 95% CI [-.04, -.20]), Irritability (β = -.22, 95% CI [-.12, -.32]), and OB factors (β = -.23, 95% CI [-.13, -.33]). Restrictiveness was estimated to have insignificant positive relations with the General factor (β = .04, 95% CI [-.06, .14]), and small positive relations with the Irritability (β = .17, 95% CI [.07, .27]), and OB factors (β = .22, 95% CI [.12, .32]).

Nurturance and Restrictiveness: Age invariance The fit of the invariant model showed no change in model fit by ΔCFI (+.003) ΔTLI (+.003), and ΔRMSEA (-.001) (Table 18). In addition, this invariant model had excellent fit by CFI (.923), TLI (.926), and RMSEA (.028). Therefore, invariance of the linear paths of Nurturance and Restrictiveness on the General, Irritability, and OB factors of ODD across offspring age could not be rejected. The standardized path coefficients are listed in Table 19. Overall, the fully age invariant model with all broad parenting dimensions included was estimated to explain 7% of the variance in the General factor (95% CI [3, 11]), and 1% and 5% of the variance in the Irritability (R^2 as % = 1, 95% CI [-1, 3]) and OB factors (R^2 as % = 5, 95% CI [-.9, 11]), respectively (Table 19).

Nurturance was estimated to have small negative relations with the General factor (β = -.22, 95% CI [-.12, -.30]), minimal negative relations with the Irritability factor (β = -.07, 95% CI [-.17, .03]), and small negative relations with the OB factor (β = -.14, 95% CI [-.04, -.24]). Restrictiveness was estimated to have small positive relations with the General factor (β = .13, 95% CI [.05, .21]), minimal positive relations with the Irritability factor (β = .08, 95% CI [-.02, .18]), and small positive relations with the OB factor (β = .18, 95% CI [.06, .30]).

Specific Parenting Practices Predicting ODD.

Monitoring, Direct Supervision, and Involvement: Sex invariance. The fit of the invariant model showed no change in model fit by ΔCFI (+.006) ΔTLI (+.006), and ΔRMSEA (-.003) (Table 18). In addition, this invariant model had excellent fit by CFI (.971), TLI (.972), and RMSEA (.029). Therefore, invariance of the linear paths of Monitoring, Direct Supervision, and Involvement on the General, Irritability, and OB factors of ODD across offspring sex could not be rejected. The standardized path coefficients are listed in Table 19. Overall, the fully sex invariant model with all specific parenting practices included was estimated to explain 8% of the variance in the General factor (95% CI [4, 12]), 4% of the variance in the Irritability factor (95% CI [.08, 8]), and 10% of the variance in the OB factor (95% CI [4, 16]) (Table 19).

Monitoring was estimated to have minimal positive relations with the General (β = .07, 95% CI [-.03, .17]), Irritability (β = .06, 95% CI [-.04, .16]), and OB factors (β = .00, 95% CI [-.10, .10]). Direct Supervision was estimated to have minimal positive relations with the General (β = .08, 95% CI [-.02, .18]) and Irritability factors (β = .09, 95% CI [-.008, .19]), and small positive relations with the OB factor (β = .22, 95% CI [.12, .32]). Involvement was estimated to have small negative relations with the General (β = -.29, 95% CI [-.21, -.37]) and Irritability factors (β = -.21, 95% CI [-.11, -.31]), and moderate negative relations with the OB factor (β = -.30, 95% CI [-.20, -.40]).

Monitoring, Direct Supervision, and Involvement: Age invariance. The fit of the invariant model showed no change in model fit by ΔCFI (.000) ΔTLI (.000), and ΔRMSEA (.000) (Table 18). In addition, this invariant model had adequate fit by CFI (.945) and TLI (.946), and excellent fit by RMSEA (.039). Therefore, invariance of the linear paths of Monitoring, Direct Supervision, and Involvement on the General, Irritability, and OB factors of ODD across offspring age could not be rejected. The standardized path coefficients are listed in Table 19. Overall, the fully age invariant model with all specific parenting practices included was estimated to explain 15% of the variance in the General factor (95% CI [9, 21]), and minimal variance in the Irritability (R^2 as % = .8, 95% CI [-.8, 2.4]) and OB factor (R^2 as % = 5, 95% CI [-.9, 11]) (Table 19).

Monitoring was estimated to have minimal positive relations with the General (β = .07, 95% CI [-.008, .15]) and Irritability factors (β = .05, 95% CI [-.05, .15]), and minimal negative relations with the OB factor (β = -.02, 95% CI [-.14, .10]). Direct Supervision was estimated to have small positive relations with the General (β = .11, 95% CI [.03, .19]) and OB factors (β = .17, 95% CI [.07, .27]), and minimal negative relations with the Irritability factor (β = -.07, 95% CI [-.17, .03]). Involvement was estimated to have moderate negative relations with the General factor (β = -.40, 95% CI [-.32, -.48]), minimal positive relations with the Irritability factor (β = .05, 95% CI [-.05, .15]), and small negative relations with the OB factor (β = -.20, 95% CI [-.08, -.32]).

All Parenting Variables Predicting ODD.

All parenting variables: Sex invariance.

The fit of the factor correlation invariant model showed no meaningful change in model fit by ΔCFI (+.004) ΔTLI (+.004), and ΔRMSEA (-.001) (Table 18). The addition of path coefficient invariance across offspring sex created no meaningful change in model fit by ΔCFI (+.004) ΔTLI (+.004), and ΔRMSEA (-.001). This final sex invariant model had adequate fit by CFI (.930) and TLI (.931), and excellent fit by RMSEA (.021). Therefore, sex invariance of all factor correlations as well as the linear paths of Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement on the General, Irritability, and OB factors of ODD could not be rejected. The factor correlations and standardized path coefficients are listed in Table 19. Overall, the fully sex invariant model with all broad and specific parenting variables included was estimated to explain 8% of the variance in the General (95% CI [2, 14]) and Irritability factors (95% CI [2, 14]), and 14% of the variance in the OB factor (95% CI [6, 22]). This did not represent a statistically significant improvement in model R^2s for the ODD factors over either the broad parenting dimensions model or the specific parenting practices model.

Correlations were estimated to be low between Nurturance and Monitoring ($r = .21$), Nurturance and Direct Supervision ($r = .13$), Restrictiveness and Monitoring ($r = .10$), Restrictiveness and Direct Supervision ($r = .16$), and Restrictiveness and Involvement ($r = .04$). The correlation between Nurturance and Involvement was estimated to be high ($r = .77$).

Nurturance was estimated to have non-significant relations with the General ($\beta = .14$, 95% CI [-.02, .30]), Irritability ($\beta = -.16$, 95% CI [-.34, .02]), and OB factor ($\beta = -.03$, 95% CI [-.21, .15]). The change in the effect of Nurturance on the General factor was statistically significant, but given sample size, the number of model parameters, and the number of statistical tests, this could easily have arisen due to chance. Restrictiveness was estimated to have non-significant relations with the General factor ($\beta = .03$, 95% CI [-.06, .14]), and small positive relations with the Irritability ($\beta = .16$, 95% CI [.06, .26]) and OB factors ($\beta = .20$, 95% CI [.10, .30]).

Monitoring was estimated to have non-significant relations with the General ($\beta = .05$, 95% CI [-.05, .15]), Irritability ($\beta = .07$, 95% CI [-.03, .17]), and OB factors ($\beta = -.01$, 95% CI [-.11, .09]). Direct Supervision was estimated to have non-significant relations with the General ($\beta = .07$, 95% CI [-.03, .17]) and Irritability factors ($\beta = .06$, 95% CI [-.06, .18]), and small positive relations with the OB factor ($\beta = .17$, 95% CI [.05, .29]). Involvement was estimated to have moderate negative relations with the General factor ($\beta = -.38$, 95% CI [-.22, -.54]), non-significant relations with the Irritability factor (β = -.10, 95% CI [-.28, .08]), and small negative relations with the OB factor (β = -.28, 95% CI [-.10, -.46]).

All parenting variables: Age invariance. The fit of the factor correlation invariant model showed no change in model fit by ΔCFI (+.002) ΔTLI (+.003), and ΔRMSEA (.000) (Table 18). The addition of invariance for the path coefficients across offspring age created no change in model fit by ΔCFI (.000) ΔTLI (-.001), and ΔRMSEA (.000). This final age invariant model had adequate fit by CFI (.913) and TLI (.914), and excellent fit by RMSEA (.023). Therefore, invariance of all factor correlations as well as the linear paths of Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement on the General, Irritability, and OB factors of ODD across offspring age could not be rejected. The factor correlations and standardized path coefficients are listed in Table 19. Overall, the fully age invariant model with all broad and specific parenting variables included was estimated to explain 16% of the variance in the General factor (95% CI [10, 22]), 4% of the variance in the Irritability factor (95% CI [.08, 8]), and 8% of the variance in the OB factor (95% CI [2, 14]). This did not represent a statistically significant improvement in model R^2s for the ODD factors over either the broad parenting dimensions model or the specific parenting practices model.

Correlations were estimated to be low between Nurturance and Monitoring (r =.25), Nurturance and Direct Supervision (r = .15), Restrictiveness and Monitoring (r = .12), Restrictiveness and Direct Supervision (r = .10), and Restrictiveness and Involvement (r = .02). The correlation between Nurturance and Involvement was estimated to be high (r = .76).

Nurturance was estimated to have insignificant relations with the General (β = .04, 95% CI [-.16, .24]) and OB factors (β = -.04, 95% CI [-.24, .16]), and small negative relations with the Irritability factor (β = -.26, 95% CI [-.08, -.44]). Restrictiveness was estimated to have small positive relations with the General (β = .11, 95% CI [.03, .19]) and OB factors (β = .18, 95% CI [.06, .30]), and insignificant relations with the Irritability factor (β = .07, 95% CI [-.05, .19]).

Monitoring was estimated to have insignificant relations with the General (β = .05, 95% CI [-.05, .15]), Irritability (β = .07, 95% CI [-.03, .17]), and OB factors (β = -.04, 95% CI [-.16, .08]). Direct Supervision was estimated to have small positive relations with the General (β = .10, 95% CI [.02, .18]) and OB factors (β = .15, 95% CI [.05, .25]), and insignificant relations with the Irritability factor (β = -.09, 95% CI [-.19, .008]). Involvement was estimated to have moderate negative relations with the General factor (β = -.47, 95% CI [-.33, -.61]), small positive relations with the Irritability factor (β = .25, 95% CI [.07, .43]), and insignificant relations with the OB factor (β = -.16, 95% CI [-.38, .06]).

Broad Parenting Dimensions Predicting CD.

Prior to incorporating any parenting influences on CD, the effects of sex (coded 0 = male, 1 = female) and age were examined to provide baseline R^2 estimates for Aggression and Rule-breaking. Sex (β = -.02, 95% CI [-.14, .10]) and age (β = .40, 95% CI [.28, .52]) were estimated to account for 16% of the variance in Rule-breaking (95% CI [6, 26]) (Table 20). Sex (β = -.19, 95% CI [-.09, -.29]) and age (β = -.33, 95% CI [-.23, -.43]) were estimated to account for 14% of the variance in Aggression (95% CI [6, 22]). The addition of Nurturance, and Restrictiveness linear paths on Aggression and Rule-breaking increased the variance accounted for in both Rule-breaking (R^2 as % = 28%, 95% CI [18, 38]) and Aggression (R^2 as % = 20%, 95% CI [12, 28]). These changes in R^2 for Aggression and Rule-breaking were non-significant. This model had excellent approximate fit by RMSEA (.029) and near-adequate approximate fit by CFI (.893) and TLI (.885) (Table 20).

Age (β = .42, 95% CI [.30, .54]) and Nurturance (β = -.32, 95% CI [-.22, -.42]) explained the bulk of the variance in Rule-breaking, with moderate paths. Sex (β = -.01, 95% CI [-.13, 11]) and Restrictiveness (β = .06, 95% CI [-.06, .18]) were estimated to have insignificant paths. Model explanation of Aggression was more evenly spread across all variables: sex (β = -.19, 95% CI [-.09, -.29]), age (β = -.31, 95% CI [-.19, -.43]), Nurturance (β = -.20, 95% CI [-.10, -.30]), and Restrictiveness (β = .17, 95% CI [.07, .27]).

Specific Parenting Practices Predicting CD.

The full sample model including sex, age, Monitoring, Direct Supervision, and Involvement linear paths on Aggression and Rule-breaking had excellent approximate fit by RMSEA (.041), and adequate approximate fit by CFI (.913) and TLI (.900). With Monitoring, Direct Supervision, and Involvement added to the baseline sex and age model, the model was estimated to account for 48% of the variance in Rule-breaking (95% CI [36, 60]) and 23% of the variance in Aggression (95% CI [15, 31]). This represents a statistically significant increase in R^2 for Rule-breaking and a non-significant increase in R^2 for Aggression over the baseline sex and age model estimates.

Age (β = .46, 95% CI [.36, .56]), Involvement (β = -.32, 95% CI [-.22, -.42]), and Direct Supervision (β = -.33, 95% CI [-.21, -.45]) explained the bulk of variance in Rule-breaking, with moderate paths. Sex (β = -.02, 95% CI [-.14, .10]) and Monitoring (β = .02, 95% CI [-.10, .14]) were estimated to have insignificant paths. Model explanation of Aggression was spread across sex (β = -.18, 95% CI [-.08, -.28]), age (β = -.26, 95% CI [-.14, -.38]), and Involvement (β = -.38, 95% CI [-.30, -.46]), with small-to-moderate paths. The lower bound estimate of the Monitoring path was barely above zero (β = .10, 95% CI [.002, .20]), and Direct Supervision was estimated to have an insignificant path (β = .06, 95% CI [-.06, .18]).

All Parenting Variables Predicting CD

The full sample model including sex, age, Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement linear paths on Aggression and Rule-breaking had excellent approximate fit by RMSEA (.024), and near-adequate approximate fit by CFI (.898) and TLI (.893). Overall, the model was estimated to account for 44% of the variance in Rule-breaking (95% CI [32, 56]) and 25% of the variance in Aggression (95% CI [17, 33]). This did not represent a statistically significant change in R^2 for Rule-breaking or Aggression over the sex, age, and broad parenting dimensions model. The R^2 for Aggression did not significantly improve beyond its estimate in the baseline sex and age model.

Age (β = .46, 95% CI [.36, .56]), Involvement (β = -.31, 95% CI [-.15, -.47]), and Direct Supervision (β = -.25, 95% CI [-.13, -.37]) explained the bulk of the variance in Rule-breaking, with small-to-moderate paths. Sex (β = -.02, 95% CI [-.14, .10]), Nurturance (β = .02, 95% CI [-.14, .18]), and Restrictiveness (β = -.02, 95% CI [-.06, .18]) were estimated to have insignificant paths, and the lower bound of the Monitoring path was barely above zero (β = -.11, 95% CI [-.01, -.21]). Model explanation of

Aggression was spread across sex (β = -.17, 95% CI [-.07, -.27]), age (β = -.23, 95% CI [-.11, -.35]), Restrictiveness (β = .17, 95% CI [.07, .27]), and Involvement (β = -.43, 95% CI [-.29, -.57]), with small-to-moderate paths. Nurturance (β = .06, 95% CI [-.08, .20]), Monitoring (β = .05, 95% CI [-.05, .15]), and Direct Supervision (β = .06, 95% CI [-.06, .18]) were estimated to have insignificant paths.

General Discussion

The primary goal of the current study was to examine the assumption of measurement invariance across sex and age for two widely used assessments of self-reported parenting as well as two parent-report DSM-IV symptom rating scales for ODD and CD. Although each of these four measures had been used to examine a variety of questions related to the etiology and course of youth disruptive behavior, none had been examined for measurement invariance across sex and age and two of the scales had never been factor analyzed. In addition, structural invariance across sex and age was unknown for all measures, as well as the sex and age invariance of the relations between the parenting and DBD measures. The general lack of evidence for sex and age invariance cast a shadow over interpretations of prior research relying on these measures, especially since the parenting and psychopathology literature easily produces instances where invariance may be violated (e.g., Butler, 2013). At minimum, the assumptions of configural, weak, and strong measurement invariance needed to be met for valid between-groups and cross-time comparisons, as well as the valid use of heterogeneous samples (Meredith, 1993; Millsap, 2011). Strict invariance was also important to consider, since a good chunk of prior research using these measures has relied on scale summary scores calculated with raw data (Meredith, 1993; Millsap, 2011). Given the number of hypotheses tested, the implications of findings for each measure will be discussed below in four separate sections, followed by two sections covering the relations between the parenting measures and each disruptive behavior measure.

Parenting scales.

Modified Child-Rearing Practices Report. The current study replicates prior work demonstrating the ability of the M-CRPR to produce two orthogonal factors representing Nurturance and Restrictiveness (Rickel and Biasatti, 1982; Deković et al., 1991). It was not expected that an essentially orthogonal two-factor structure would represent the best-fitting model based on the RMSEA.LB index, which has been shown in a recent simulation study to outperform other commonly used fit indices such as the RMSEA, AIC, and BIC in recovering the "true" factor structure when EFA-based methods are employed (Preacher et al., 2013). Given the large number of items, it was expected that the "true" factor structure as identified by the RMSEA.LB index would be more multidimensional. This may reflect the fact that the items of the M-CRPR were chosen based on an orthogonal PCA analysis of the original 91-item Q-Sort CRPR precisely because they tapped into the broad and stable dimensions of Nurturance and Restrictiveness (Rickel & Biasatti, 1982).

The Nurturance factor contained 19 items asking about parental warmth and affection, preference for the use of positive reinforcement and inductive reasoning, sensitivity to child needs, encouragement of child autonomy, and enjoyment of the parental role and parent-child relationship (Table 2). The Restrictiveness factor contained 27 items asking about the level of strict control asserted over child behaviors, beliefs, and emotional expressions, application of performance pressure, discouragement of autonomy, and preference for the use of criticism, physical discipline, and guilt induction (Table 2).

Aside from replicating this orthogonal two-factor model, another goal of the current study was to examine whether an ESEM model would offer advantages over traditional CFA models. Although low factor correlations in both the CFA and ESEM versions of this model did not confer the advantage of better discriminant validity to the ESEM model, the ESEM model did fit significantly better to the underlying data due to significant item cross-loadings. These cross-loadings support the imperfect nature of the items of the M-CRPR as measures of single constructs, an important assumption of CFA (Asparouhov & Muthén, 2009; Marsh et al., 2014).

The primary goal of the study related to the M-CRPR was to investigate the measurement and structural invariance of its factor structure across sex and age. Using a multigroup ESEM approach, the M-CRPR was found to have full measurement and structural invariance across sex, and near complete measurement and structural invariance across age. Only one item asking parents about the extent to which they forbade “talking badly” about a teacher showed non-invariance, with parents demonstrating lower thresholds to report applying this restriction to offspring expression if their offspring was a child. Since the M-CRPR is a 46-item scale, incorporating this single instance of threshold non-invariance (and the resulting non-invariance of item 13 residual variance), is unlikely to adversely affect the utility of the measurement model (Byrne et al., 1989).

Supervision/Involvement Scale. An important goal of the current study related to the S/I scale was to explore its factor structure, as it has become a popular measure in various influential longitudinal studies of the etiology of youth psychopathology and antisocial behavior (Loeber et al., 1998; Loeber et al., 2000; Keenan et al., 2010), but has never been factor analyzed. Test developers designed the S/I scale to capture two important parenting constructs based on a meta-analysis of the family correlates of delinquency – Supervision and Involvement (Loeber et al., 1998). Developers hypothesized that Supervision could be divided into two sub-scales - *Poor Supervision* and *No Set Time Home* - and that Involvement could be divided into four sub-scales - *Low Family Talk*, *Low Family Activities*, *Boy Not Involved*, and *Don't Enjoy Boy* (Loeber et al., 1998). The current analysis supported Loeber and colleagues' (1998) broad Supervision (a.k.a. Monitoring) and Involvement factors, and found evidence for an additional Direct Supervision factor consisting of previously unused items.

Monitoring consisted of six items focused on the process by which parents track the activities, associates, and whereabouts of their child when they are separated from the child. This is consistent with the most widely accepted definition of parental monitoring (Dishion & McMahon, 1998). A popular argument in recent years has been that questionnaires tapping parental monitoring naturally divide monitoring into control (i.e., rules parents use to facilitate the monitoring process), solicitation (i.e., questions parents ask their children to gain information about their time away from parental oversight), and knowledge (i.e., parents' perceived knowledge of their child's time away from parental oversight, Stattin & Kerr, 2000). These arguments have been based on confirmatory CFA models (Racz & McMahon, 2011). This pattern was not observed for the S/I scale using

ESEM, although the S/I scale items were not designed with this three-factor conceptualization of parental monitoring in mind (Loeber et al., 1998).

Direct Supervision consisted of five items focused on the extent to which parents ensure their children remain in close proximity to and under the watchful eyes of a caregiver. No prior studies have made use of the items preferentially loading on Direct Supervision. Involvement consisted of 12 items focused on the general level and quality of parent-child engagement in shared activities as well as the general level and quality of parent-child communication.

These factors had excellent discriminant validity as represented by their low factor correlations. As expected, significant item cross-loadings were common and there was a small contingent of items that seemed to discriminate poorly between these three inferred constructs, offering information that could theoretically apply to some or all of the constructs. Although these items might have to be cast aside for the scale to be useful in generating manifest summary scores, in a latent ESEM approach, they could be retained and all available information could be utilized. Also as expected, model fit worsened significantly when item cross-loadings were restricted to zero in a CFA version of the model, reflecting the fallible nature of the items as indicators of single constructs. In addition, factor correlations were approximately doubled in the CFA model, reducing model discriminant validity (ESEM: .16, .20, .30; CFA: .36, .47, .46, respectively).

The primary goal of this study related to the S/I scale was to examine its measurement and structural invariance across sex and age. It was predicted that at least factor mean differences would be apparent based on offspring age for the original Supervision and Involvement constructs (Loeber et al., 1998), given prior longitudinal research relying on random samplings of adolescents' daily interactions with family (Larson et al., 1996). As predicted, only factor means demonstrated non-invariance across offspring age, with parents of adolescents reporting more Monitoring (*d* = .52), less Direct Supervision (*d* = -1.42), and less Involvement (*d* = -.31) than parents of children. These mean differences are consistent with the normative changes in the parent-child relationship during adolescence reported by Larson and colleagues (1996). Monitoring, a communicative process between parent and child focused heavily on what the child does when they are out of parental view (Dishion & McMahon, 1998; Hayes, Hudson, & Matthews, 2007), increased with offspring age. Direct Supervision, the amount of time spent in close proximity to parents and under their direct oversight, decreased with offspring age. Involvement, a construct strongly influenced by time spent in shared activities with parents, decreased with offspring age, but to a lesser degree than Direct Supervision, likely due to the aspects of Involvement focused on the general level and quality of parent-child communication.

These findings are supportive of the prior use of the S/I scale in samples where sex or age varied (Loeber et al., 1998; Lahey et al., 1999; Burke, Loeber, Lahey, & Rathouz, 2005; Burke, Pardini, & Loeber, 2008; Pardini et al., 2008; Hipwell et al., 2010; Scott et al., 2013). If the factor structure identified in the current study replicates consistently across sufficiently powered independent samples, future researchers can use these findings as a guide for selecting the item make-up of manifest scales.

Disruptive Behavior Disorder Scales.

Emory Diagnostic Rating Scale – ODD. An important goal of the current study related to the ODD scale was to compare the CFA modified bifactor model identified in prior research using the GTS sample data and four other large, diverse, and independent samples (Burke et al., 2014) to an ESEM-derived model. In particular, we were keen to examine whether freely estimated item cross-loadings provided a better representation of the high level of shared variance among the items than an orthogonal general factor, and better discriminant validity in the form of lower factor correlations. Despite the allowance of item cross-loadings in the selected two-factor ESEM model, little meaningful improvement in factor discriminant validity in the form of lower factor correlations was observed (ESEM: r = .72; CFA: average r = ~.80; Burke et al., 2014). When this empirically-derived model was compared to the a priori CFA model, model fit actually appeared to (slightly) favor the CFA modified bifactor model. Given its greater parsimony and good track record of cross-validation (Burke et al., 2014), the CFA modified bifactor model remains the better approximation of the DSM-IV ODD factor structure. The problem of poor discriminant validity for the Irritability and OB specific factors in this modified bifactor model remains unresolved.

The primary goal of the current study related to the ODD scale was to examine its measurement and structural invariance across sex and age. As predicted, the modified bifactor model was found to have near complete measurement and structural invariance across sex, and near complete measurement and structural invariance across age. Only item 6, “touchy or easily annoyed by others,” in the child-adolescent comparison and item 2, “argues with adults,” in both the child-adolescent and male-female comparisons demonstrated non-invariance in item residual variance. Items 2 and 6 had lower residual variance if parents of an adolescent were responding to the questionnaire. The same was true for parent responses to item 2 if their offspring was female. Inspection of standardized residual variance estimates for item 2 revealed that these differences may have little substantive impact on sex or age comparisons. These differences may be more meaningful for age comparisons using item 6, although the lower bound of the 95% confidence interval for the estimated residual variance in the child group was barely above the fixed value of zero in the adolescent group. These instances of item residual non-invariance would have no impact on the utility of the EDRS ODD scale when latent variable models are used. However, potential item residual non-invariance might affect the utility of the scale if manifest scale scores were needed since the scale consists of only 8 items.

Emory Diagnostic Rating Scale – CD. An important goal of the current study related to the CD rating scale was to replicate the two-factor Aggression and Rule-breaking structure found in prior research (e.g., Tackett et al., 2003) and to see whether using ESEM over CFA would improve the discriminant validity of this model. Results clearly favored an oblique Rule-breaking and Aggression ESEM model that largely conformed to previous EFA models (e.g., Tackett et al., 2003) and had superior discriminant validity relative to a two-factor CFA model (factor correlations: ESEM $r = .42$; ICM-CFA $r = .94$). This supports prior assertions that Rule-breaking and Aggression are distinct dimensions of CD based on differences in etiology (Burt, 2009; Burt & Klump, 2012; Burt, 2013), developmental trajectories and rank-order stability (Tremblay, 2010), prevalence across sex (Moffitt, 2003), and external correlates (Burt, 2012; Lahey & Waldman, 2012). Rule-breaking consisted of two items tapping truancy and curfew violations. Aggression consisted of six items tapping the frequency of lying to get others in trouble or to get out of trouble, use of bullying, threats, and physical cruelty, and starting fights in and out of the home. The remaining three items contributed information to both factors (lying to get out of responsibilities, lying to get one's way, and destruction of property). Only 11 items had sufficient variation to be included in all planned factor analytic and path analyses, forcing decisions about the "true" internal structure of DSM-IV CD symptoms to be based on only 7 out of 15 symptoms. Items lacking in variability were the same items estimated to have low population prevalence in prior large, nationally representative epidemiological surveys relying on self-report (i.e., sexual assault (.3%), theft with confrontation (1.5%), arson (1.8%), use of a weapon (3.4%), physical cruelty towards animals (4.2%), breaking and entering (6.4%), running away (12.7%), and theft without confrontation (~15%, Nock, Kazdin, Hiripi, & Kessler, 2006). This low base-rate problem is a major limiting factor for the usefulness of factor analysis in examining the internal structure of CD.

The primary goal of the current study related to the EDRS CD scale was to examine its measurement and structural invariance across sex and age. Unfortunately, only four to five items had sufficient variation for multigroup invariance testing, necessitating the use of MIMIC modeling in the full sample to complete limited tests of invariance. Results revealed the plausibility of item threshold invariance across sex and age (only linear paths were tested) and non-invariance of factor means. This adds to the very limited research supporting item threshold invariance across sex for DSM-IV CD symptoms (Meier et al., 2009). No prior studies have examined item threshold invariance across age using DSM symptoms. In terms of mean differences, parents reported .19 (95% CI [-.09, -.29]) standard deviations less Aggression in female offspring than male offspring and a one standard deviation increase in age was estimated to correspond to a

.33 (95% CI [-.23, -.43]) standard deviation decline in Aggression. In contrast, a one standard deviation increase in age was estimated to correspond to a .40 (95% CI [.28, .52]) standard deviation increase in Rule-breaking behavior. Sex differences in Rule-breaking could not be distinguished from zero (β = -.02, 95% CI [-.14, .10]).

These age-based findings are consistent with previous reviews demonstrating the distinctive trajectories of Aggression and Rule-breaking across development (Tremblay, 2010). These findings suggest that the frequency of Aggression peaks around age 2-4, and then declines sharply while maintaining remarkable rank-order stability, with the most aggressive individuals at age 2-4 typically remaining the most aggressive individuals at later ages (Tremblay, 2010). Rule-breaking follows the opposite pattern, increasing from late childhood to its zenith in adolescence, with greater instability in the rank ordering of individuals (Tremblay, 2010). A greater mean level of Aggression and Rule-breaking among males has been a consistent finding in the research literature (Moffitt, 2003), but only sex differences in Aggression were replicated in the current study. Further work is needed to clarify the source of this inconsistency across studies.

Unfortunately, the MIMIC modeling approach employed was unable to test the crucial assumptions of configural and weak invariance. These assumptions must be satisfied in order for these item threshold and factor mean invariance findings to be valid (Meredith, 1993; Millsap & Yun-Tein, 2004; Millsap, 2011). Therefore, the results of this study regarding CD measurement and structural invariance should not be over-interpreted.

Parenting as a Predictor of the DBDs.

Predicting ODD. The next goal of the current study was to examine the multigroup sex and age prediction invariance of all parenting constructs related to ODD. The correlations among all parenting variables, as well as the incremental validity of specific practices over broad dimensions of parenting were also of interest.

The sex and age invariance of the correlations between all parenting factors could not be rejected. As predicted, correlations between Restrictiveness (psychological control) and parenting practices likely falling along the dimension of behavioral control (Direct Supervision and Monitoring) were low. Similarly, correlations between Nurturance (responsiveness) and Direct Supervision and Monitoring were low. These low correlations were expected given the putatively orthogonal nature of responsiveness and demandingness (Maccoby & Martin, 1983; Baumrind, 1991), and psychological control and behavioral control (Barber et al., 1994). Also as predicted, Involvement, a parenting practice likely saturated by responsiveness via its emphasis on the level and quality of parent-child communication and reciprocal engagement in shared activities, had large positive correlations with Nurturance and insignificant correlations with Restrictiveness. While it is likely that the magnitude of the correlations between Nurturance and Involvement would be lower if their correlations had been estimated in an ESEM model rather than an ESEM-within-CFA model, the relative magnitude of the Nurturance and Involvement correlations would still be large. This was confirmed in a supplementary five-factor oblique ESEM analysis treating all parenting items as if they had come from one scale. Not only did this analysis recover the factor structure of each scale when they were examined separately and produce excellent model fit statistics, but the correlation between Nurturance and Involvement (r = .41) was still around double all other factor correlations (results available upon request).

In all sex and age multigroup prediction invariance analyses related to ODD, the invariance of all path coefficients could not be rejected. This invariance across sex and age supports prior research that has used the M-CRPR (Atlas & Rickel, 1988; Holden et al., 1995; Stormont-Spurgin & Zentall, 1995; Gonzales et al., 1996; Mason et al., 1996) and the S/I scale (Loeber et al., 1998; Lahey et al., 1999; Pardini et al., 2008; Hipwell et al., 2010; Scott et al., 2013) to examine the relations between parenting and DBD symptoms in samples of mixed sex or age.

Nurturance and Restrictiveness were estimated to account for a small amount of variance in the ODD General factor in both the sex (95% CI [.04, 4]) and age invariance models (95% CI [3, 11]) (Table 19), but this was not the case for the Irritability or OB factors (Table 19). The addition of Monitoring, Direct Supervision, and Involvement increased the variance explained for all ODD factors, such that all 95% confidence interval lower bound estimates of R^2 were above 0%. However, with only a single exception, none of these increases constituted a statistically significant change. This was not strong evidence in favor of the incremental validity of the specific parenting practices. On their own, Monitoring, Direct Supervision, and Involvement were unable to explain more than 0% variance in any ODD factor, except for the General factor (sex: 95% CI [4, 12]; age: 95% CI [9, 21]). In the final models with all parenting constructs included, only Involvement had consistently significant path coefficients for the ODD General factor (sex: 95% CI [-.22, -.54]; age 95% CI [-.33, -.61]), only Restrictiveness (sex: 95% CI [.10, .30]; age: 95% CI [.06, .30]) and Direct Supervision (sex: 95% CI

[.05, .29]; age: 95% CI [.05, .25]) had consistently significant path coefficients for the OB factor, and no variable had consistently significant path coefficients for the Irritability factor. Overall, moderate associations were found between increases in the level and quality of parent-child communication and engagement in shared activities and decreases in the General factor underlying all ODD symptoms. Conversely, increasing levels of a harsh and psychologically controlling approach to child-rearing as well as increasing direct oversight of- and close proximity to- the child had small associations with increases in the aspects of the oppositional behavior symptoms unrelated to the General factor. Theoretical and empirical work suggests that these relations between parenting and ODD may be bidirectional in nature, with aversive behaviors in the parent-child dyad escalating over time and negatively reinforcing children for acting coercively and parents for "giving in" (e.g., Reid et al., 2002; Kazdin, 2008). These simple cross-sectional associations are consistent with this literature, including interventions that improve child oppositionality via reductions in parental coercion and inconsistency and increases in parental positive involvement (e.g., Kaminski et al., 2008; Forgatch et al., 2009). It is potentially interesting that child Irritability did not seem to be strongly related to the parenting variables investigated, and future experimental studies might consider whether an underlying disposition to experience negative affect is less plastic in the face of parenting behaviors than overt opposition (although it should be remembered that the correlation between Irritability and OB is excessively high and the full range of possible parenting behaviors was not covered in the current study).

Predicting CD. Although low variation in item responses precluded the use of multigroup prediction invariance analyses, and insufficient sample size precluded testing the moderation of parenting relations by sex and age as well as synergistic relations among the parenting variables, the simple linear paths of sex, age, and parenting on CD could still be estimated. In addition, the secondary objective of testing the incremental validity of the specific parenting practices over and above the broad dimensions of parenting could still be addressed at the level of linear relations. In terms of 95% confidence interval estimates, sex and age alone were estimated to explain 6-26% of the variance in Rule-breaking and 6-22% of the variance in Aggression. The addition of Nurturance and Restrictiveness generated non-significant increases in variance explained (18-38% in Rule-breaking and 12-28% for Aggression). In contrast, when Monitoring, Direct Supervision, and Involvement were added to the sex and age baseline model, variance explained increased significantly for Rule-breaking (36-60%) and non-significantly for Aggression (15-31%). With nearly identical point estimates and standard errors, it appears that Direct Supervision and Involvement jointly drove this statistically significant improvement in R^2 for Rule-breaking. The addition of Nurturance and Restrictiveness to this sex, age, and specific parenting practices model generated no more statistically significant improvements in model R^2s. As predicted, the more behaviorally specific and proximal parenting practices had more robust relations with CD than the broad dimensions of parenting.

The final model contained the statistically significant relations of age (95% CI [.36, .56]), Monitoring (95% CI [-.01, -.21]), Direct Supervision (95% CI [-.13, -.37]), and Involvement (95% CI [-.15, -.47]) with Rule-breaking, and the statistically significant relations of sex (95% CI [-.07, -.27]), age (95% CI [-.11, -.35]), Restrictiveness (95% CI [.07, .27]), and Involvement (95% CI [-.29, -.57]) with

Aggression. This suggests that increases in the tracking of child behavior, associates, and whereabouts when the child is away from caregivers, increases in direct oversight of-, and physical proximity to the child, and increases in the level and quality of parent-child communication and engagement in shared activities, are all associated with decreases in the Rule-breaking symptoms of CD, controlling for sex and age. These results also suggest that increases in the level and quality of parent-child communication and engagement in shared activities, and a general approach to childrearing that is less harsh and psychologically controlling, is associated with decreases in Aggression, controlling for sex and age. The apparently stronger negative relation between parental oversight and Rule-Breaking makes sense given the popular interpretation of Rule-Breaking as consisting of covert forms of antisocial behavior, whereas Aggression consists of overt forms and is more strongly related to oppositionality (e.g., Frick et al., 1993). Given the number of missing Rule-breaking and Aggression-relevant items, it is unknown how well these effect sizes would generalize to a sample with sufficient item response variation.

These findings for ODD and CD are consistent with prior research demonstrating negative relationships between Involvement and conduct problems during childhood and adolescence among boys (Pardini et al., 2008), precocious sexual behavior among early adolescent girls (Hipwell et al., 2010), and depression and CD severity during adolescence among girls (Scott et al., 2013). They are also consistent with prior work showing positive relations between Restrictiveness and spanking in early childhood (Holden et al., 1995), aggression in preschool (Stormont-Spurgin & Zentall, 1995), antisocial behavior from early-to-mid-adolescence (Mason et al., 1996), and poor school achievement among adolescents in low-risk neighborhoods (Gonzales et al., 1996). While monitoring has been shown to have small-to-moderate relations with child delinquency in prior meta-analyses (e.g., Hoeve et al., 2009), in the current study, weak and insignificant relations were observed. In contrast, Direct Supervision appeared to have more consistently significant relations with child disruptive behavior. It is no surprise that parents might have more power to influence child behavior when they are able to directly observe child behavior, rather than relying on child self-report. As expected, general declines in the Nurturance path coefficients were observed when Involvement was included due to redundancy. The Involvement paths were robust to this redundancy, consistent with the hypothesis that more proximal and behaviorally specific parenting variables have stronger relations with child outcomes (Darling & Steinberg, 1993). The small-to-moderate effect size range for all parenting relations with disruptive behavior was expected based on prior meta-analyses of correlational (e.g., Hoeve et al., 2009) and intervention studies (e.g., Kaminski et al., 2008).

Strengths & Limitations.

Strengths of this sample included its size (sub-group range: N = 447-535; full sample range: N = 940-1,131) and representativeness of the general population of Georgia in the early 1990s based on race, ethnicity, socioeconomic status, and household location (Waldman et al. 1998). Strengths of this analysis included the flexible application of CFA and ESEM methodology for determining the optimal factor structure of each scale, the application of multigroup methodology for the determination of measurement, structural, and prediction invariance across sex and age for three out of four scales, and reliance on the WLSMV estimator for categorical data rather than treating the items as continuous normal. Factor analyses were strengthened by comparing alternative a priori and empirically-derived measurement models using statistical indices of relative and approximate model fit. Invariance analyses were strengthened by examining all levels of invariance using statistical indices of relative and approximate model fit in three out of four scales. SEM analyses were strengthened by the use of ESEM measurement models where appropriate, which reduced non-essential multicollinearity and model misspecification, especially for the EDRS CD Scale.

Limitations of this sample include the low variation in CD symptoms and some of the items on the S/I Scale. This issue could be resolved by oversampling for at-risk or clinic-referred subjects. In general, reliance on a community-based twin sample and a mailed survey sampling strategy based on state birth records may have restricted the range of reported parenting and DBD symptoms. In addition, an even larger sample would have also allowed multigroup invariance to be examined across the conjunction of sex *and* age using four sex-age groups, rather than separate multigroup analyses defined by sex *or* age. A sex *and* age multigroup analysis would have provided more information regarding whether the measures tested can be validly used in youth samples that vary in sex *and* age, which is typically the case. In addition, an even larger sample would have allowed synergistic effects among the parenting variables to be examined. The existence of synergistic effects, such as the notion that the effectiveness of specific practices depends on the affective context created by broad dimensions of parenting (Darling & Steinberg, 1993), is an important question in the parenting literature. Another limitation of the current analysis that could have been improved upon with an even larger sample size was the dichotomization of age in the multigroup invariance analyses, which reduced statistical power. This limitation is inherent to any multigroup invariance analysis that uses continuous grouping variables, and while some multigroup/MIMIC hybrid models have been recently reported in the literature to mitigate this problem, even these more advanced modeling procedures cannot fully account for the loss of information inherent to this approach (Marsh et al., 2013).

Aside from improvements that could be made with an even larger sample with better representation at the extremes of reported parenting and child behavior, two major issues of study design were shared method variance and cross-sectional measurement. Shared method variance created by asking parents to report on both the quality of their parenting and the adjustment outcomes of their offspring may have created inflations in the relations between parenting and child adjustment, as well as other biases in parameter estimates that are difficult to predict (Bank, Dishion, Skinner, & Patterson, 1990). Such undesirable method effects could be corrected using a multi-informant-multi-method study design, in which parents, children, and other informants could have been asked to report on all outcomes (Bank et al., 1990). At the very least, different informants could have been used for the independent and dependent variables. It would have also been interesting and informative to include observational measures of parenting behavior, since parent report on their own behavior may not correspond well to their actual behavior (e.g., Reid et al., 2002).

The use of a passive, cross-sectional study design precluded our ability to make inferences about the directions of effects in the current study (Bell, 1968; Rutter, Pickles, Murray, & Eaves, 2001). This is an important limitation, as a cornerstone assumption of many prominent theories of parenting influences on DBDs in the natural environment is the assumption of bidirectional effects (e.g., Patterson, 1982). A longitudinal design would have offered information on this interesting question, and would have also allowed the assumption of measurement, structural, and prediction invariance across time to be examined. The M-CRPR and the S/I scale have been used in prior longitudinal studies (e.g., Mason et al., 1996; Loeber et al., 1998; Loeber et al., 2000; Keenan et al., 2010), but have not been tested for the invariance of psychometric properties over time.

A final limitation was also one of the strengths of this study design – use of ESEM measurement models. While operating on more “realistic” assumptions regarding the presence of item cross-loadings, these models have far more freely estimated parameters than more restrictive CFA models (Asparouhov & Muthén, 2009; Marsh et al., 2014). This higher number of freely estimated parameters increases standard errors and decreases the precision of parameter estimates, although this reduced precision should be offset by reduced bias in parameter estimates (Asparouhov & Muthén, 2009; Marsh et al., 2014). This higher number of freely estimated parameter estimates may also increase the risk of type I errors due to multiple comparisons, although this issue is offset by the reliance on statistical indices of overall model fit for making judgments between alternative models, which substantially reduces the total number of statistical tests. Regardless, no corrections for multiple comparisons were attempted, and therefore these findings await replication.

Future Directions.

Future research should use larger, more diverse longitudinal samples that oversample at-risk, adjudicated, or clinic-referred subjects in order to capture the full range of parenting and DBD symptoms and to permit invariance analyses across the conjunction of sex *and* age, as well as invariance across other key grouping variables, such as race/ethnicity, socioeconomic status, and time. Examining invariance across markers of relative disadvantage may be especially important, as current evidence suggests that the effects of Nurturance and Restrictiveness on child adjustment may depend on race/ethnicity and exposure to deviant peers (Mason et al., 1996) and the effects of parental monitoring may depend on neighborhood safety (Robinson et al., 2015). Interpretation of such interesting and important findings is limited by the current dearth of information on psychometric invariance in the parenting and psychopathology literature and beyond (Lubke et al., 2004; Borsboom, 2006; Millsap, 2007; Chen, 2008). Fortunately, large longitudinal samples oversampled for the presence of DBD symptoms and adverse child-rearing behaviors, spanning childhood and adolescence, relying on multiple informants, and making use of the S/I Scale already exist (e.g., Loeber et al., 1998; Loeber et al., 2000; Keenan et al., 2010). Therefore, the majority of the limitations of the current sample could be addressed by simply pooling data across studies. An important future direction of this work is to seek collaboration with colleagues in order to conduct more informative and conclusive invariance analyses.

Ultimately, it is impossible to simultaneously examine invariance across all possible grouping variables. Even if this were possible, this still doesn't tell us whether the psychometric properties of the scales are the same for all individuals (i.e. local homogeneity, Millsap, 2011). However, it is still valuable to confirm invariance across the conjunction of a limited set of theoretically important grouping variables, and future researchers are advised to routinely examine this assumption to prevent questions of methodological artifact from clouding the interpretation of substantive findings.

References

Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Aebi, M., Muller, U. C., Asherson, P., Banaschewski, T., Buitelaar, J., Ebstein, R., . . .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.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317-332.

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

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397-438.

Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. *Mplus Technical Appendix*, 1-10.

Atlas, J. G., & Rickel, A. U. (1988). Maternal coping styles and adjustment in children. *Journal of Primary Prevention*, *8*(4), 169-185.

Bank, L., Dishion, T., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with “glop.” *Depression and aggression in family interaction*, (pp. 247-279). Hillsdale, NJ: Erlbaum.

Barber, B. K., Olsen, J. E., & Shagle, S. C. (1994). Associations between parental psychological and behavioral control and youth internalized and externalized behaviors. *Child Development*, *65*(4), 1120–1136.

Baumrind, D. (1971). Current patterns of parental authority. *Developmental Psychology*, *4*(1-2), 1-103.

Baumrind, D. (1991). Parenting styles and adolescent development. In J. Brooks-Gunn, R. Lerner, & A. C. Petersen (Eds.), *The encyclopedia of adolescence* (pp. 746-758). New York: Garland.

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review, 75*(2), 81-95.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238-246.

Block, J. (1965). *The child rearing practices report.* Berkeley: University of California, Institute of Human Development.

Block, J. (1973). Conceptions of sex role: Some cross-cultural and longitudinal perspectives. *American Psychologist, 28*(6), 512-526.

Bonnelykke, B., Hauge, M., Holm, N., Kristoffersen, K., & Gurtler, H. (1989). Evaluation of zygosity diagnosis in twin pairs below age seven by means of a mailed questionnaire. *Acta Geneticae Medicae Gemellologiae, 38*(3-4), 305–313.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425–440.

Bridge, J. A., Goldstein, T. R., & Brent, D. A. (2006). Adolescent suicide and suicidal behavior. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 372-394.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111-150.

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

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

Burke, J. D., Hipwell, A. E., & Loeber, R. (2010a). 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. *Journal of Child Psychology and Psychiatry*, *46*(11), 1200-1210.

Burke, J. D., Pardini, D. A., & Loeber, R. (2008). Reciprocal relationships between parenting behavior and disruptive psychopathology from childhood through adolescence. *Journal of Abnormal Child Psychology*, *36*(5), 679–692.

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

Burns, G. L., Walsh, J. A., Gomez, R., & Hafetz, N. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*, *18*(4), 452-457.

Burns, G. L., Walsh, J. A., Servera, M., Lorenzo-Seva, U., Cardo, E., & Rodríguez-

Fornells, A. (2013). Construct validity of ADHD/ODD rating scales: recommendations for the evaluation of forthcoming DSM-V ADHD/ODD scales. *Journal of Abnormal Child Psychology*, *41*(1), 15-26.

Burt, S. A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. *Psychological Bulletin*, *135*(4), 608-637.

Burt, S. A. (2012). How do we optimally conceptualize the heterogeneity within antisocial behavior? An argument for aggressive versus non-aggressive behavioral dimensions. *Clinical Psychology Review*, *32*(4), 263-279.

Burt, S. A. (2013). Do etiological influences on aggression overlap with those on rule breaking? A meta-analysis. *Psychological Medicine*, *43*(9), 1801-1812.

Burt, S. A., Donnellan, M. B., Iacono, W. G., & McGue, M. (2011). Age-of-onset or behavioral sub-types? A prospective comparison of two approaches to characterizing the heterogeneity within antisocial behavior. *Journal of Abnormal Child Psychology*, *39*(5), 633-644.

Burt, S. A., & Klump, K. L. (2012). Etiological distinctions between aggressive and non-aggressive antisocial behavior: Results from a nuclear twin family model. *Journal of Abnormal Child Psychology*, *40*(7), 1059-1071.

Burt, S. A., Rescorla, L. A., Achenbach, T. M., Ivanova, M. Y., Almqvist, F., Begovac, I., ... & Verhulst, F. C. (2015). The association between aggressive and non-aggressive antisocial problems as measured with the Achenbach System of Empirically Based Assessment: A study of 27,861 parent–adolescent dyads from 25 societies. *Personality and Individual Differences*, *85*, 86-92.

Butler, A. M. (2013). Cross-racial measurement equivalence of the Eyberg Child Behavior Inventory factors among low-income young African American and Non-Latino White children. *Assessment*, *20*(4), 484-495.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466.

Caron, A., Weiss, B., Harris, V., & Catron, T. (2006). Parenting behavior dimensions and child psychopathology: Specificity, task dependency, and interactive relations. *Journal of Clinical Child and Adolescent Psychology*, *35*(1), 34-45.

Chen, F.F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005-1018.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, *29*(4), 468-508.

Cheung, G.W. & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155-159.

Cohen, M. A., & Piquero, A. R. (2009). New evidence on the monetary value of saving a high risk youth. *Journal of Quantitative Criminology*, *25*(1), 25-49.

Cote, L. R., & Azar, S. T. (1997). Child age, parent and child gender, and domain differences in parents' attributions and responses to children's outcomes. *Sex Roles*, *36*(1-2), 23-50.

Cosgrove, V. E., Rhee, S. H., Gelhorn, H. L., Boeldt, D., Corley, R. C., Ehringer, M. A., ... & Hewitt, J. K. (2011). Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *Journal of Abnormal Child Psychology*, *39*(1), 109-123.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin*, *113*(3), 487-496.

Deković, M., Janssens, J. M., & Gerris, J. R. (1991). Factor structure and construct validity of the Block Child Rearing Practices Report (CRPR). *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(2), 182-187.

Dick, D. M., Viken, R. J., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2005). 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*, *33*(2), 219-229

Dishion, T. J., & McMahon, R. J. (1998). Parental monitoring and the prevention of child and adolescent problem behavior: A conceptual and empirical formulation. *Clinical Child and Family Psychology Review, 1*(1), 61-75.

Dishion, T. J., & Patterson, G. R. (2006). The development and ecology of antisocial behavior in children and adolescents. In D. Cicchetti & D. J. Cohen

(Eds.), *Developmental psychopathology Vol. 3: Risk, disorder, and adaptation* (pp. 503-541). New York: Wiley.

Dong, L., Wu, H., & Waldman, I. D. (2014). Measurement and structural invariance of the Antisocial Process Screening Device. *Psychological Assessment*, *26*(2), 598-608.

Fergusson, D. M., Horwood, J. L., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry*, *46*(8), 837-849.

Ficks, C. A., Lahey, B. B., & Waldman, I. D. (2013). Does low birth weight share common genetic or environmental risk with childhood disruptive disorders? *Journal of Abnormal Psychology*, *122*(3), 842-853.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466-491.

Forehand, R., Lafko, N., Parent, J., & Burt, K. B. (2014). Is parenting the mediator of change in behavioral parent training for externalizing problems of youth? *Clinical Psychology Review*, *34*(8), 608-619.

Forgatch, M. S., Patterson, G. R., Degarmo, D. S., & Beldavs, Z. G. (2009). Testing the Oregon delinquency model with 9-year follow-up of the Oregon Divorce Study. *Development and Psychopathology*, *21*(02), 637-660.

Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Van Horn, Y., Christ, M. A. G., ...

& 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.

Gonzales, N. A., Cauce, A. M., Friedman, R. J., & Mason, C. A. (1996). Family, peer, and neighborhood influences on academic achievement among African-American adolescents: One-year prospective effects. *American Journal of Community Psychology*, *24*(3), 365-387.

Guay, F., Morin, A. J., Litalien, D., Valois, P., & Vallerand, R. J. (2014). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education*, *83*(1), 51-82.

Harden, K. P., Patterson, M. W., Briley, D. A., Engelhardt, L. E., Kretsch, N., Mann, F. D., ... & Tucker-Drob, E. M. (2015). Developmental changes in genetic and environmental influences on rule-breaking and aggression: age and pubertal development. *Journal of Child Psychology and Psychiatry, 56*(12), 1370-1379.

Hayes, L., Hudson, A., & Mathews, J. (2007). Understanding parental monitoring through the analysis of monitoring episodes in context. *International Journal of Behavioral Consultation and Therapy, 3*(1), 96-108.

Hipwell, A. E., Keenan, K., Loeber, R., & Battista, D. (2010). Early predictors of sexually intimate behaviors in an urban sample of young girls. *Developmental Psychology*, *46*(2), 366-378.

Hirschi, T. (1969). *Causes of delinquency*. Berkeley, CA: University of California.

Hoeve, M., Dubas, J. S., Eichelsheim, V. I., van der Laan, P. H., Smeenk, W., & Gerris,

J. R. (2009). The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*, *37*(6), 749-775.

Holden, G. W., & Edwards, L. A. (1989). Parental attitudes toward child rearing: Instruments, issues, and implications. *Psychological Bulletin*, *106*(1), 29-58.

Holden, G. W., Coleman, S. M., & Schmidt, K. L. (1995). Why 3-year-old children get spanked: Parent and child determinants as reported by college-educated mothers. *Merrill-Palmer Quarterly 41*(4), 431-452.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Jackson, R. W., Sneider, H., Davis, H., & Treiber, F. A. (2001). Determination of twin zygosity: A comparison of DNA with various questionnaire indices. *Twin Research*, *4*(1), 12–18.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika, 34*(2), 183-202.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*(351), 631-639.

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*(4), 567-589.

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin*, *102*(2), 187-203. doi:10.1037/0033-2909.102.2.187

Kazdin, A. E. (2008). *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents*. New York: Oxford University Press.

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 & Adolescent Psychology*, *39*(4), 506-521.

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

Lahey, B. B., Gordon, R. A., Loeber, R., Stouthamer-Loeber, M., & Farrington, D. P. (1999). Boys who join gangs: A prospective study of predictors of first gang entry. *Journal of Abnormal Child Psychology*, *27*(4), 261-276.

Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D., & Rathouz, P. J. (2011). Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of General Psychiatry*, *68*(2), 181-189.

Lahey, B. B., & Waldman, I. D. (2012). Annual research review: Phenotypic and causal structure of conduct disorder in the broader context of prevalent forms of psychopathology. *Journal of Child Psychology and Psychiatry*, *53*(5), 536-557.

Lambert, E. W., Wahler, R. G., Andrade, A. R., & Bickman, L. (2001). Looking for the disorder in conduct disorder. *Journal of Abnormal Psychology*, *110*(1), 110-123.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202-220.

Larson, R. W., Richards, M. H., Moneta, G., Holmbeck, G., & Duckett, E. (1996). Changes in adolescents' daily interactions with their families from ages 10 to 18: Disengagement and transformation. *Developmental Psychology*, *32*(4), 744-754.

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*, *42*(6), 937-951.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum Associates.

Loeber, R., Green, S. M., Lahey, B. B., Frick, P. J., & McBurnett, K. (2000). Findings on disruptive behavior disorders from the first decade of the Developmental Trends Study. *Clinical Child and Family Psychology Review*, *3*(1), 37-60.

Loeber, R., & Stouthamer-Loeber, M. (1986). Family factors as correlates and predictors of juvenile conduct problems and delinquency. *Crime & Justice*, *7*, 29-149.

Loehlin, J. C. (2004). *Latent variable models, 4th edition: An introduction to factor, path, and structural equation analysis*. Mahwah, NJ: Lawrence Erlbaum & Associates, Inc.

Lubke, G. H., Dolan, C. V., & Neale, M. C. (2004). Implications of absence of measurement invariance for detecting sex limitation and genotype by environment interaction. *Twin Research*, *7*(03), 292-298.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99.

Maccoby, E. E, & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interaction. In E. M. Hetherington (Ed.) & P. H. Mussen (Series Ed.), *Handbook of child psychology Vol. 4: Socialization, personality, and social development* (pp. 1-101). New York: Wiley.

Mandell, D. S., Guevara, J. P., Rostain, A. L., & Hadley, T. R. (2003). Economic grand rounds: Medical expenditures among children with psychiatric disorders in a Medicaid population. *Psychiatric Services*, *54*(4), 465-467.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85-110.

Marsh, H. W., Nagengast, B., & Morin, A. J. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology*, *49*(6), 1194-1218.

Marsh, H. W., Tracey, D. K., & Craven, R. G. (2006). Multidimensional self-concept structure for preadolescents with mild intellectual disabilities: A hybrid multigroup-MIMIC approach to factorial invariance and latent mean differences. *Educational and Psychological Measurement*, *66*(5), 795-818.

Mason, C. A., Cauce, A. M., Gonzales, N., & Hiraga, Y. (1996). Neither too sweet nor too sour: Problem peers, maternal control, and problem behavior in African American adolescents. *Child Development*, *67*(5), 2115-2130.

McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: A meta-analysis. *Journal of Abnormal Child Psychology*, *34*(4), 525-541.

McCord, J. (1979). Some child-rearing antecedents of criminal behavior in adult men. *Journal of Personality and Social Psychology*, *37*(9), 1477-1486.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1-21.

McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, *38*(1), 75-79.

McKee, L., Colletti, C., Rakow, A., Jones, D. J., & Forehand, R. (2008). Parenting and child externalizing behaviors: Are the associations specific or diffuse? *Aggression and Violent Behavior*, *13*(3), 201-215.

Meier, M. H., Slutske, W. S., Heath, A. C., & Martin, N. G. (2009). The role of harsh discipline in explaining sex differences in conduct disorder: A study of opposite-sex twin pairs. *Journal of Abnormal Child Psychology*, *37*(5), 653–664.

Meredith W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4) 525–543.

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research*, *30*(4), 577-605.

Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika*, *72*(4), 461-473.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York:

Routledge.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479-515.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674-701.

Moffitt, T. E. (2003). Life-course-persistent and adolescence-limited antisocial behavior: A 10-year research review and a research agenda. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.), *Causes of conduct disorder and juvenile delinquency*, (pp. 49-75). New York: Guilford Press.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*(4), 557-585.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide, 7th Edition*. Los Angeles, CA: Authors.

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological Medicine*, *36*(5), 699-710.

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication. *Journal of Child Psychology and Psychiatry*, *48*(7), 703-713.

O'Connor, T. G. (2002). Annotation: The 'effects' of parenting reconsidered: Findings, challenges, and applications. *Journal of Child Psychology and Psychiatry*, *43*(5), 555-572.

Pardini, D. A., Fite, P. J., & Burke, J. D. (2008). Bidirectional associations between parenting practices and conduct problems in boys from childhood to adolescence: The moderating effect of age and African-American ethnicity. *Journal of Abnormal Child Psychology*, *36*(5), 647–662.

Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28-56.

Racz, S. J., & McMahon, R. J. (2011). The relationship between parental knowledge and monitoring and child and adolescent conduct problems: A 10-year update. *Clinical Child and Family Psychology Review*, *14*(4), 377-398.

Reid, J. B., Patterson, G. R., & Snyder, J. E. (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention*. Washington, D.C.: American Psychological Association.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Rickel, A. U., & Biasatti, L. L. (1982). Modification of the block child rearing practices report. *Journal of Clinical Psychology*, *38*(1), 129-134.

Robinson, B. A., Winiarski, D. A., Brennan, P. A., Foster, S. L., Cunningham, P. B., & Whitmore, E. A. (2015). Social context, parental monitoring, and multisystemic therapy outcomes. *Psychotherapy*, *52*(1), 103-110.

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.

Rutter, M., Pickles, A., Murray, R., & Eaves, L. (2001). Testing hypotheses on specific environmental causal effects on behavior. *Psychological Bulletin*, *127*(3), 291-324.

Sampson, R. J., & Laub, J. H. (1994). Urban poverty and the family context of delinquency: A new look at structure and process in a classic study. *Child Development*, *65*(2), 523-540.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Scott, L. N., Whalen, D. J., Zalewski, M., Beeney, J. E., Pilkonis, P. A., Hipwell, A. E., & Stepp, S. D. (2013). Predictors and consequences of developmental changes in adolescent girls' self-reported quality of attachment to their primary caregiver. *Journal of Adolescence*, *36*(5), 797–806.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.

Singh, A. L., & Waldman, I. D. (2010). The etiology of associations between negative emotionality and childhood externalizing disorders. *Journal of Abnormal Psychology*, *119*(2), 376-388.

Smetana, J. G., Campione-Barr, N., & Metzger, A. (2006). Adolescent development in interpersonal and societal contexts. *Annual Review of Psychology*, *57*, 255-284.

Spitz, E., Moutier, R., Reed, T., Busnel, M. C., Marchaland, C., Roubertoux, P. L., &

Carlier, M. (1996). Comparative diagnoses of twin zygosity by SSLP variant analysis, questionnaire, and dermatoglyphic analysis. *Behavior Genetics*, *26*(1), 55–63.

Stattin, H., & Kerr, M. (2000). Parental monitoring: A reinterpretation. *Child Development*, *71*(4), 1072-1085.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180.

Stormont-Spurgin, M., & Zentall, S. S. (1995). Contributing factors in the manifestation of aggression in preschoolers with hyperactivity. *Journal of Child Psychology and Psychiatry*, *36*(3), 491-509.

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

Stringaris, A., Zavos, H., Leibenluft, E., Maughan, B., & Eley, T. C. (2012). Adolescent irritability: Phenotypic associations and genetic links with depressed mood. *The American Journal of Psychiatry*, *169*(1), 47–54.

Tackett, J. L., Krueger, R. F., Iacono, W. G., & McGue, M. (2005). Symptom-based subfactors of *DSM*-defined conduct disorder: Evidence for etiologic distinctions. *Journal of Abnormal Psychology*, *114*(3), 483–487.

Tackett, J. L., Krueger, R. F., Sawyer, M. G., & Graetz, B. W. (2003). Subfactors of DSM-IV conduct disorder: Evidence and connections with syndromes from the Child Behavior Checklist. *Journal of Abnormal Child Psychology*, *31*(6), 647-654.

Ten Haaf, P. G. J., Janssens, J. M. A. M., & Gerris, J. R. (1994). Child-rearing measures:

Convergent and discriminant validity. *European Journal of Psychological Assessment*, *10*(2), 111-128.

Tremblay, R. E. (2010). Developmental origins of disruptive behaviour problems: The ‘original sin’ hypothesis, epigenetics and their consequences for prevention. *Journal of Child Psychology and Psychiatry*, *51*(4), 341-367.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.

Tuvblad, C., Zheng, M., Raine, A., & Baker, L. A. (2009). 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*, *37*(2), 153-167.

Waldman, I. D., Rhee, S. H., Levy, F., & Hay, D. A. (2001). Causes of the overlap among symptoms of attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. In D.A. Hay & F. Levy (Eds.), *Attention, Genes and ADHD*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Waldman, I.D., Rowe, D.C., Abramowitz, A., Kozel, S.T., Mohr, J.H., Sherman, S.L,. . .Stever, C. (1998). Association and linkage of the dopamine transporter gene and attention deficit hyperactivity disorder in children: Heterogeneity owing to diagnostic subtype and severity. *American Journal of Human Genetics*, *63*(6), 1767-1776.

Watkins, M. W. (2002). Rc: Coefficient of Congruence [Computer software]. Phoenix, AZ: Ed & Psych Associates, Inc.

West, D. J., & Farrington, D. P. (1973). *Who becomes delinquent? Second report of the Cambridge Study in Delinquent Development.* London: Heinemann Educational.

Yarnell, L. M., Sargeant, M. N., Prescott, C. A., Tilley, J. L., Farver, J. A. M., Mednick, S. A.,. . .Luczak, S. E. (2013). Measurement invariance of internalizing and externalizing behavioral syndrome factors in a non-western sample. *Assessment*, *20*(5), 642-655.

Table 1
Modified Child-Rearing Practices Report: Exploratory ESEM goodness-of-fit statistics

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	ΔCFI	ΔTLI	$\Delta RMSEA$
8r Full	1040(695)***	.960	.940	.021 [.018, .024]	__	__	__	__
8r Boy	827(695)***	.964	.946	.020 [.014, .025]	__	__	__	__
8r Girl	881(695)***	.960	.941	.022 [.018, .027]	__	__	__	__
8r Child	797(695)**	.969	.954	.017 [.010, .022]	__	__	__	__
8r Teen	845(695)***	.965	.947	.021 [.015, .026]	__	__	__	__
7r Full	1187(734)***	.947	.926	.023 [.021, .026]	154(39)***	-0.013	-0.014	0.002
7r Boy	907(734)***	.952	.933	.022 [.017, .026]	88(39)***	-0.012	-0.013	0.002
7r Girl	982(734)***	.947	.926	.025 [.021, .029]	108(39)***	-0.013	-0.015	0.003
7r Child	871(734)***	.959	.942	.019 [.013, .024]	78(39)***	-0.01	-0.012	0.002
7r Teen	929(734)***	.954	.935	.023 [.018, .027]	92(39)***	-0.011	-0.012	0.002
6r Full	1324(774)***	.936	.914	.025 [.023, .027]	147(40)***	-0.011	-0.012	0.002
6r Boy	989(774)***	.941	.921	.024 [.019, .028]	93(40)***	-0.011	-0.012	0.002
6r Girl	1084(774)***	.934	.912	.027 [.023, .031]	116(40)***	-0.013	-0.014	0.002
6r Child	958(774)***	.945	.926	.021 [.016, .026]	95(40)***	-0.014	-0.016	0.002
6r Teen	1005(774)***	.945	.927	.024 [.020, .029]	89(40)***	-0.009	-0.008	0.001
5r Full	1477(815)***	.923	.902	.027 [.025, .029]	171(41)***	-0.013	-0.012	0.002
5r Boy	1090(815)***	.924	.904	.026 [.022, .030]	112(41)***	-0.017	-0.017	0.002
5r Girl	1187(815)***	.921	.899	.029 [.026, .033]	116(41)***	-0.013	-0.013	0.002
5r Child	1054(815)***	.928	.909	.024 [.019, .028]	103(41)***	-0.017	-0.017	0.003
5r Teen	1125(815)***	.927	.907	.028 [.024, .031]	129(41)***	-0.018	-0.02	0.004
4r Full	1659(857)***	.907	.887	.029 [.027, .031]	196(42)***	-0.016	-0.015	0.002
4r Boy	1191(857)***	.908	.889	.028 [.024, .032]	117(42)***	-0.016	-0.015	0.002
4r Girl	1278(857)***	.910	.892	.030 [.027, .034]	112(42)***	-0.011	-0.007	0.001
4r Child	1161(857)***	.909	.890	.026 [.022, .030]	115(42)***	-0.019	-0.019	0.002
4r Teen	1238(857)***	.910	.891	.030 [.026, .033]	127(42)***	-0.017	-0.016	0.002
3r Full	1860(900)***	.888	.871	.031 [.029, .033]	214(43)***	-0.019	-0.016	0.002
3r Boy	1309(900)***	.887	.871	.030 [.027, .034]	135(43)***	-0.021	-0.018	0.002
3r Girl	1393(900)***	.895	.879	.032 [.029, .035]	132(43)***	-0.015	-0.013	0.002

3r Child	1270(900)***	.889	.872	.028 [.024, .031]	124(43)***	-0.02	-0.018	0.002
3r Teen	1341(900)***	.896	.880	.031 [.028, .035]	122(43)***	-0.014	-0.011	0.001
2r Full	**2589(944)*****	**.808**	**.790**	**.039 [.037, .041]**	**520(44)*****	**-0.08**	**-0.081**	**0.008**
2r Boy	**1651(944)*****	**.806**	**.787**	**.039 [.036, .042]**	**279(44)*****	**-0.081**	**-0.084**	**0.009**
2r Girl	**1742(944)*****	**.830**	**.814**	**.040 [.037, .043]**	**284(44)*****	**-0.065**	**-0.065**	**0.008**
2r Child	**1560(944)*****	**.815**	**.797**	**.035 [.032, .038]**	**249(44)*****	**-0.074**	**-0.075**	**0.007**
2r Teen	**1585(944)*****	**.849**	**.834**	**.037 [.034, .040]**	**226(44)*****	**-0.047**	**-0.046**	**0.006**
1 Full	5977(989)***	.419	.392	.067 [.065, .068]	1384(45)***	-0.389	-0.398	0.028
1 Boy	3066(989)***	.429	.402	.065 [.063, .068]	694(45)***	-0.377	-0.385	0.026
1 Girl	3105(989)***	.550	.529	.063 [.061, .066]	663(45)***	-0.28	-0.285	0.023
1 Child	2763(989)***	.466	.442	.058 [.056, .061]	592(45)***	-0.349	-0.355	0.023
1 Teen	3041(989)***	.515	.493	.064 [.062, .067]	670(45)***	-0.334	-0.341	0.027
Sample Size: Full = 1,131; Boy = 495 ; Girl = 534 ; Child = 528 ; Teen = 499								

Note. ¶p < .10; * p < .05; ** p < .01; *** p < .001; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; r = correlated factors; bold font indicates best model by fit, factor discriminant validity, interpretability, consistency across sex and age sub-samples, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; Δ□χ^2 = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 2

Modified Child-Rearing Practices Report: Initial measurement models

	Standardized factor loadings											
	Nurturance						*Restrictiveness*					
	CFA	*ESEM*					*CFA*	*ESEM*				
Items	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*
1. I respect my child's opinions and encourage her/him to express them	.64*	**.64***	**.62***	**.68***	**.62***	**.71***		-.06	-.15*	-.01	-.01	-.07
2. I don't think young children of different sexes should be allowed to see each other naked		-.04	-.11*	.01	.06	-.13	.18*	**.19***	**.20***	**.12***	**.12**	**.19***
3. I feel a child should be given comfort and understanding when s/he is scared or upset	.65*	**.66***	**.73***	**.63***	**.56***	**.74***		.02	-.02	.05	.13	-.06
4. I try to keep my child away from children or families who have different ideas or values from our own		.03	.09	-.06	.04	-.01	.31*	**.30***	**.35***	**.22***	**.23***	**.32***
5. I believe physical punishment to be the best way of disciplining		-.16*	-.22*	-.12	-.16*	-.19	.39*	**.40***	**.38***	**.40***	**.36***	**.40***
6. I believe that a child should be seen and not heard		-.18*	-.19*	-.20	-.27*	-.15	.46*	**.47***	**.41***	**.48***	**.33***	**.57***
7. I express affection by hugging, kissing, and holding my child	.63*	**.64***	**.60***	**.70***	**.75***	**.59***		-.10*	-.07	-.11	-.19*	-.03
8. I find some of my greatest satisfactions in my child	.58*	**.58***	**.62***	**.55***	**.51***	**.64***		.10*	.14*	.02	-.04	.19*
9. I prefer that my child not try things if there is a chance s/he will fail		-.16*	-.13*	-.19*	-.21*	-.13	.35*	**.36***	**.36***	**.35***	**.26***	**.44***
10. I encourage my child to wonder and think about life	.61*	**.61***	**.58***	**.66***	**.67***	**.60***		-.09*	-.07	-.08	-.05	-.05
11. I usually take into account my child's preferences in making plans for the family	.50*	**.49***	**.45***	**.54***	**.53***	**.49***		-.01	-.01	.00	.01	.03
12. I feel a child should have time to think, daydream, and even loaf sometimes	.55*	**.55***	**.48***	**.61***	**.56***	**.56***		-.10*	-.10	-.04	-.01	-.10
13. I do not allow my child to say bad things about her/his teachers		.09*	.08	.11	**.19***	.00	.23*	**.23***	**.24***	**.17***	.14*	**.25***
14. I teach my child that in one way or another punishment will find him/her if s/he misbehaves		.17*	.08	.21*	.13*	.16	.45*	**.44***	**.49***	**.38***	**.50***	**.36***
15. I do not allow my child to get angry with me		-.15*	-.14*	-.16	-.18*	-.15	.41*	**.42***	**.42***	**.42***	**.41***	**.41***
16. I am easygoing and relaxed with my child	.49*	**.48***	**.45***	**.49***	**.39***	**.53***		-.08*	-.07	-.03	-.08	.01

17. I talk it over and reason with my child when s/he misbehaves	.55*	**.54***	**.53***	**.54***	**.50***	**.58***		.01	.05	.01	.12*	.01
18. I trust my child to behave as s/he should, even when I am not with her/him	.49*	**.51***	**.49***	**.53***	**.33***	**.64***		.06	.08	.02	.06	.09
19. I joke and play with my child	.66*	**.66***	**.58***	**.72***	**.65***	**.68***		-.05	.06	-.12*	.00	-.03
20. My child and I have warm, intimate times together	.71*	**.70***	**.69***	**.72***	**.76***	**.67***		-.07*	-.05	-.07	-.07	-.02
21. I encourage my child to be curious, to explore and question things	.77*	**.77***	**.73***	**.83***	**.81***	**.76***		-.06	-.09*	-.02	.02	-.08
22. I expect my child to be grateful and appreciate all the advantages s/he has		.39*	.39*	.40*	.30*	**.45***	.47*	**.45***	**.48***	**.41***	**.46***	.43*
23. I believe in toilet training a child as soon as possible		-.02	-.09	.05	-.01	-.04	.47*	**.48***	**.50***	**.46***	**.45***	**.49***
24 - I believe in praising a child when s/he is good and think it gets better results than punishing her/him when s/he is bad	.34*	**.33***	**.42***	**.26***	**.28***	**.38***		.01	.06	-.01	.06	.01
25. I make sure my child knows that I appreciate what s/he tries or accomplishes	.74*	**.74***	**.71***	**.76***	**.66***	**.77***		.06	.09	.05	.17*	.01
26. I encourage my child to talk about his/her troubles	.75*	**.75***	**.69***	**.81***	**.68***	**.79***		.06	.01	.09	.16*	-.01
27. I believe children should not have secrets from parents		.02	.07	.00	.02	.02	.41*	**.41***	**.36***	**.44***	**.36***	**.42***
28. I teach my child to keep control of her/his feelings at all times		.05	.10*	.02	.04	.06	.50*	**.50***	**.51***	**.46***	**.46***	**.51***
29 - I dread answering my child's questions about sex		-.17*	-.19*	-.13	-.13*	-.21*	.28*	**.29***	**.26***	**.30***	**.25***	**.27***
30. When I am angry with my child, I let her/him know it		.20*	.15*	.26*	.09	**.27***	.37*	**.36***	**.33***	**.39***	**.48***	**.27***
31. I think a child should be encouraged to do things better than others		.01	.03	-.02	-.08	.07	.54*	**.54***	**.54***	**.54***	**.55***	**.56***
32. I believe that scolding and criticism make my child improve		-.19*	-.19*	-.17	-.09	-.27*	.55*	**.55***	**.55***	**.54***	**.59***	**.51***
33. I believe my child should be aware of how much I sacrifice for her/him		-.08*	-.08	-.07	-.05	-.12	.64*	**.64***	**.70***	**.59***	**.63***	**.67***
34. I do not allow my child to question my decisions		-.12*	-.09	-.15	-.17*	-.12	.38*	**.38***	**.29***	**.51***	**.35***	**.44***
35. I let my child know how ashamed and disappointed I am when s/he misbehaves		.07	.04	.10	-.03	.12	.58*	**.58***	**.55***	**.58***	**.61***	**.51***

36. I want my child to be independent of me	.24*	**.25***	**.22***	**.30***	.15*	**.36***		.13*	.12*	.12*	**.29***	.00
37. I find it interesting and educational to be with my child for long periods	.61*	**.60***	**.60***	**.64***	**.55***	**.68***		.14*	.19*	.07	.17*	.16*
38. I instruct my child not to get dirty while s/he is playing		-.19*	-.24*	-.18	-.24*	-.22	.38*	**.39***	**.41***	**.44***	**.44***	**.39***
39. I control my child by warning her/him about the bad things that can happen to her/him		-.08*	-.10	-.08	-.11	-.09	.51*	**.51***	**.53***	**.52***	**.50***	**.54***
40. I think it is best if the mother, rather than the father, is the one with the most authority over the children		-.11*	-.05	-.15	-.08	-.15	.35*	**.35***	**.40***	**.37***	**.38***	**.37***
41. I don't want my child to be looked upon as different than others		.00	.00	-.01	.05	-.07	.36*	**.36***	**.37***	**.30***	**.28***	**.37***
42. I don't think children should be given sexual information before they can understand everything		-.03	-.04	-.02	.00	-.07	.31*	**.32***	**.36***	**.22***	**.29***	**.25***
43. I encourage my child always to do her/his best	.40*	**.42***	**.35***	**.50***	.29*	**.51***		.38*	.34*	.40*	**.45***	.31*
44. I think it is good practice for a child to perform in front of others		.23*	.28*	.20*	.21*	.24*	.36*	**.35***	**.40***	**.31***	**.42***	**.34***
45. I expect a great deal of my child		.21*	.21*	.20*	.08	.30*	.47*	**.47***	**.50***	**.47***	**.45***	**.59***
46. I feel that it is good for a child to play competitive games		.26*	.26*	.27*	.19*	.31*	.38*	**.37***	**.43***	**.30***	**.43***	**.32***
ω	.91	.91	.90	.92	.89	.92	.85	.85	.85	.84	.84	.85
α	.82	.82	.80	.85	.80	.85	.80	.80	.81	.79	.79	.81
Rickel & Biasatti (1982) r_c		.98	.97	.98	.97	.98		.97	.97	.96	.95	.97

Factor Correlations

	CFA	*ESEM*				
Factor pairs	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*
Nurturance and Restrictiveness	.03	.01	.04	.005	.03	-.001

Sample Size: Full = 1,131; Boy = 495 ; Girl = 534 ; Child = 528 ; Teen = 499

Note. $^*p$ < .05; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; bolded loadings = highest loading for each item within the full-sample or a sub-sample; boxes mark the factor that each item preferentially loads upon in ESEM; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006); r_c = coefficient of congruence relative to standardized factor loadings for 18-item Nurturance and 20-item Restrictiveness factors reported in Rickel & Biasatti (1982) using PCA with Varimax rotation (.98-1.00 = excellent, .92-.98 = good, .82-.92 = borderline, .68-.82 = poor; MacCallum, Widaman, Zhang, & Hong, 1999).

Table 3

Modified Child-Rearing Practices Report best model comparisons: ESEM vs. CFA

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
ESEM Full	**2589(944)*****	**.808**	**.790**	**.039 [.037, .041]**	—	—	—	—
CFA Full	3261(988)***	.735	.723	.045 [.043, .047]	443(44)***	-0.073	-0.067	0.006
ESEM Male	**1651(944)*****	**.806**	**.787**	**.039 [.036, .042]**	—	—	—	—
CFA Male	2017(988)***	.717	.704	.046 [.043, .049]	264(44)***	-0.089	-0.083	0.007
ESEM Female	**1742(944)*****	**.830**	**.814**	**.040 [.037, .043]**	—	—	—	—
CFA Female	2068(988)***	.770	.759	.045 [.043, .048]	245(44)***	-0.06	-0.055	0.005
ESEM Child	**1560(944)*****	**.815**	**.797**	**.035 [.032, .038]**	—	—	—	—
CFA Child	1819(988)***	.750	.738	.040 [.037, .043]	206(44)***	-0.065	-0.059	0.005
ESEM Teen	**1585(944)*****	**.849**	**.834**	**.037 [.034, .040]**	—	—	—	—
CFA Teen	2110(988)***	.735	.722	.048 [.045, .051]	308(44)***	-0.114	-0.112	0.011
Sample Size: Full = 1,131; Boy = 495 ; Girl = 534 ; Child = 528 ; Teen = 499								

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; bold font indicates best model by fit, factor discriminant validity, interpretability, consistency across sex and age sub-samples, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 4

Modified Child-Rearing Practices Report: Measurement and structural invariance across age and sex

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
Sex								
Configural	3395(1888)***	.817	.799	.039 [.037, .042]	__	__	__	__
Weak	3224(1976)***	.848	.841	.035 [.033, .037]	88(88)	0.031	0.042	-0.004
Strong	3319(2105)***	.852	.855	.033 [.031, .036]	171(129)*	0.004	0.014	-0.002
Strict	3328(2151)***	.857	.862	.033 [.030, .035]	59(46)	0.005	0.007	.000
Factor Variance-Covariance	3215(2154)***	.871	.876	.031 [.029, .033]	5(3)	0.014	0.014	-0.002
Factor Means	**3195(2156)*****	**.874**	**.879**	**.031 [.028, .033]**	**0.3(2)**	**0.003**	**0.003**	**.000**
Age								
Configural	3157(1888)***	.831	.815	.036 [.034, .038]	__	__	__	__
Weak	3079(1976)***	.853	.846	.033 [.031, .035]	121(88)	0.022	0.031	-0.003
Strong	3308(2103)***	.839	.842	.033 [.031, .036]	330(127)*	-0.014	-0.004	0.000
Partial Strong[a]	3260(2100)***	.845	.848	.033[.031, .035]	265(124)***	-0.008	0.002	.000
Strict[a]	3327(2146)***	.843	.848	.033 [.031, .035]	120(46)*	-0.002	0.000	.000
Factor Variance-Covariance[a]	3252(2149)***	.853	.858	.032 [.029, .034]	8(3)	0.01	0.01	-0.001
Factor Means[a]	**3244(2151)*****	**.854**	**.860**	**.031 [.029, .034]**	**3(2)**	**0.001**	**0.002**	**-0.001**
Sample Size: Boy = 495 ; Girl = 534 ; Child = 528 ; Teen = 499								

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; bold font indicates maximum level of invariance that cannot be rejected; a = item 13 thresholds and uniqueness non-invariant; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 5

Modified Child-Rearing Practices Report: Final ESEM measurement models

	Standardized Factor Loadings			
	Nurturance		*Restrictiveness*	
	All invariant	*All invariant (-item 13 τ, δ)*	*All Invariant*	*All invariant (-item 13 τ, δ)*
Items	*Boy Girl*	*Child Teen*	*Boy Girl*	*Child Teen*
1. I respect my child's opinions and encourage her/him to express them	**.66***	**.66***	-.08*	-.06
2. I don't think young children of different sexes should be allowed to see each other naked	-.05	-.05	**.17***	**.16***
3. I feel a child should be given comfort and understanding when s/he is scared or upset	**.67***	**.67***	.02	.02
4. I try to keep my child away from children or families who have different ideas or values from our own	.02	.01	**.28***	**.28***
5. I believe physical punishment to be the best way of disciplining	-.17*	-.17*	**.38***	**.38***
6. I believe that a child should be seen and not heard	-.19*	-.21*	**.44***	**.46***
7. I express affection by hugging, kissing, and holding my child	**.65***	**.65***	-.09	-.12*
8. I find some of my greatest satisfactions in my child	**.58***	**.59***	.08	.08
9. I prefer that my child not try things if there is a chance s/he will fail	-.16*	-.17*	**.36***	**.36***
10. I encourage my child to wonder and think about life	**.63***	**.63***	-.08*	-.06
11. I usually take into account my child's preferences in making plans for the family	**.50***	**.51***	.00	.00
12. I feel a child should have time to think, daydream, and even loaf sometimes	**.56***	**.56***	-.07*	-.07
13. I do not allow my child to say bad things about her/his teachers	.10*	.08*	**.20***	**.19***
14. I teach my child that in one way or another punishment will find him/her if s/he misbehaves	.16*	.15*	**.43***	**.43***
15. I do not allow my child to get angry with me	-.15*	-.16*	**.42***	**.41***
16. I am easygoing and relaxed with my child	**.47***	**.47***	-.05	-.04
17. I talk it over and reason with my child when s/he misbehaves	**.54***	**.55***	.03	.06
18. I trust my child to behave as s/he should, even when I am not with her/him	**.51***	**.52***	.06	.06
19. I joke and play with my child	**.66***	**.66***	-.03	-.04

20. My child and I have warm, intimate times together	**.70***	**.71***	-.05	-.08*
21. I encourage my child to be curious, to explore and question things	**.79***	**.79***	-.06	-.05
22. I expect my child to be grateful and appreciate all the advantages s/he has	.40*	.40*	**.45***	**.45***
23. I believe in toilet training a child as soon as possible	-.02	-.03	**.48***	**.47***
24 - I believe in praising a child when s/he is good and think it gets better results than punishing her/him when s/he is bad	**.34***	**.33***	.03	.03
25. I make sure my child knows that I appreciate what s/he tries or accomplishes	**.75***	**.73***	.06	.06
26. I encourage my child to talk about his/her troubles	**.76***	**.75***	.05	.05
27. I believe children should not have secrets from parents	.04	.02	**.40***	**.40***
28. I teach my child to keep control of her/his feelings at all times	.06	.05	**.48***	**.49***
29 - I dread answering my child's questions about sex	-.16*	-.17*	**.28***	**.26***
30. When I am angry with my child, I let her/him know it	.20*	.21*	**.36***	**.36***
31. I think a child should be encouraged to do things better than others	.00	.00	**.54***	**.56***
32. I believe that scolding and criticism make my child improve	-.18*	-.19*	**.55***	**.55***
33. I believe my child should be aware of how much I sacrifice for her/him	-.07*	-.07*	**.65***	**.65***
34. I do not allow my child to question my decisions	-.13*	-.14*	**.41***	**.39***
35. I let my child know how ashamed and disappointed I am when s/he misbehaves	.07	.06	**.56***	**.57***
36. I want my child to be independent of me	**.26***	**.28***	.12*	.14*
37. I find it interesting and educational to be with my child for long periods	**.63***	**.62***	.13*	.15*
38. I instruct my child not to get dirty while s/he is playing	-.21*	-.22*	**.43***	**.41***
39. I control my child by warning her/him about the bad things that can happen to her/him	-.09*	-.09*	**.53***	**.52***
40. I think it is best if the mother, rather than the father, is the one with the most authority over the children	-.11*	-.11*	**.38***	**.38***
41. I don't want my child to be looked upon as different than others	.00	-.01	**.34***	**.33***
42. I don't think children should be given sexual information before they can understand everything	-.03	-.03	**.29***	**.27***
43. I encourage my child always to do her/his best	**.43***	**.44***	.37*	.37*
44. I think it is good practice for a child to perform in front of others	.24*	.24*	**.35***	**.37***
45. I expect a great deal of my child	.21*	.21*	**.49***	**.51***
46. I feel that it is good for a child to play competitive games	.27*	.26*	**.36***	**.37***

ω	.91	.91	.85	.85
α	.83	.83	.80	.80
Rickel & Biasatti (1982) r_c	.99		.99	

Factor Correlations				
Factor pairs	*Boy*	*Girl*	*Child*	*Teen*
Nurturance and Restrictiveness	.03		.02	
Sample Size: Boy = 495 ; Girl = 534 ; Child = 528 ; Teen = 499				

Note. $^*p$ < .05; Boy = male group; Girl = female group; Child = child group; Teen = adolescent group; boxes mark the factor that each item preferentially loads upon in ESEM; highest loading for each item is bolded for each multigroup model; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006); r_c = coefficient of congruence relative to standardized factor loadings for 18-item Nurturance and 20-item Restrictiveness factors reported in Rickel & Biasatti (1982) using PCA with Varimax rotation (.98-1.00 = excellent, .92-.98 = good, .82-.92 = borderline, .68-.82 = poor; MacCallum, Widaman, Zhang, & Hong, 1999).

Table 6

Supervision/Involvement Scale: Exploratory ESEM of goodness-of-fit statistics

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
8r Full	376(202)***	.986	.972	.028 [.023, .032]	__	__	__	__
8r Boy	267(202)**	.988	.976	.026 [.16, .033]	__	__	__	__
8r Girl	257(202)**	.993	.986	.023 [.013, .030]	__	__	__	__
8r Child	250(202)*	.986	.972	.021 [.011, .029]	__	__	__	__
8r Teen	246(202)*	.995	.989	.021 [.009, .029]	__	__	__	__
7r Full	469(224)***	.980	.964	.031 [.027, .035]	91(22)***	-0.006	-0.008	+0.003
7r Boy	328(224)***	.981	.965	.031 [.023, .038]	62(22)***	-0.007	-0.011	0.005
7r Girl	298(224)***	.990	.983	.025 [.017, .032]	45(22)**	-0.003	-0.003	0.002
7r Child	294(224)**	.980	.964	.024 [.016, .032]	45(22)**	-0.006	-0.008	0.003
7r Teen	295(224)**	.991	.984	.025 [.017, .033]	51(22)***	-0.004	-0.005	0.004
6r Full	584(247)***	.973	.955	.035 [.031, .038]	121(23)***	-0.007	-0.009	+0.004
6r Boy	390(247)***	.974	.957	.034 [.028, .041]	64(23)***	-0.007	-0.008	0.003
6r Girl	373(247)***	.984	.973	.031 [.024, .037]	73(23)***	-0.006	-0.01	0.006
6r Child	343(247)***	.973	.955	.027 [.020, .034]	52(23)***	-0.007	-0.009	0.003
6r Teen	415(247)***	.979	.965	.037 [.031, .043]	108(23)***	-0.012	-0.019	0.012
5r Full	793(271)***	.958	.937	.041 [.038, .045]	185(24)***	-0.015	-0.019	+0.006
5r Boy	491(271)***	.960	.939	.040 [.035, .046]	104(24)***	-0.014	-0.018	0.006
5r Girl	469(271)***	.974	.962	.037 [.031, .043]	98(24)***	-0.01	-0.011	0.006
5r Child	411(271)***	.960	.940	.031 [.025, .037]	71(24)***	-0.013	-0.015	0.004
5r Teen	515(271)***	.969	.954	.043 [.037, .048]	97(24)***	-0.01	-0.011	0.006
4r Full	1068(296)***	.938	.915	.048 [.045, .051]	237(25)***	-0.02	-0.022	+0.007
4r Boy	601(296)***	.944	.923	.046 [.040, .051]	107(25)***	-0.016	-0.016	0.006
4r Girl	637(296)***	.956	.939	.047 [.042, .051]	153(25)***	-0.018	-0.023	0.01
4r Child	508(296)***	.940	.917	.037 [.031, .042]	100(25)***	-0.02	-0.023	0.006
4r Teen	645(296)***	.956	.940	.049 [.043, .054]	126(25)***	-0.013	-0.014	0.006
3r Full	**1369(322)*** **	**.916**	**.894**	**.054 [.051, .057]**	**276(26)*** **	**-0.022**	**-0.021**	**+0.006**
3r Boy	**729(322)*** **	**.925**	**.906**	**.051 [.046, .055]**	**130(26)*** **	**-0.019**	**-0.017**	**0.005**
3r Girl	**815(322)*** **	**.936**	**.920**	**.054 [.049, .058]**	**172(26)*** **	**-0.02**	**-0.019**	**0.007**
3r Child	**646(322)*** **	**.908**	**.884**	**.044 [.039, .049]**	**136(26)*** **	**-0.032**	**-0.033**	**0.007**
3r Teen	**849(322)*** **	**.934**	**.917**	**.057 [.053, .062]**	**183(26)*** **	**-0.022**	**-0.023**	**0.008**

2r Full	2602(349)***	.819	.789	.076 [.073, .078]	724(27)***	-0.097	-0.105	+0.022
2r Boy	1337(349)***	.818	.789	.076 [.071, .080]	378(27)***	-0.107	-0.117	0.025
2r Girl	1384(349)***	.866	.844	.075 [.070, .079]	388(27)***	-0.07	-0.076	0.021
2r Child	1032(349)***	.806	.774	.061 [.057, .065]	320(27)***	-0.102	-0.11	0.017
2r Teen	1322(349)***	.878	.858	.075 [.070, .079]	338(27)***	-0.056	-0.059	0.018
1 Full	4637(377)***	.657	.631	.100 [.097, .103]	1264(28)***	-0.162	-0.158	+0.024
1 Boy	2260(377)***	.654	.627	.100 [.096, .104]	615(28)***	-0.164	-0.162	0.024
1 Girl	2428(377)***	.734	.714	.101 [.097, .105]	782(28)***	-0.132	-0.13	0.026
1 Child	1814(377)***	.591	.560	.085 [.081, .089]	572(28)***	-0.215	-0.214	0.024
1 Teen	2179(377)***	.774	.756	.098 [.094, .102]	557(28)***	-0.104	-0.102	0.023
Sample Size: Full = 1,130; Boy = 495; Girl = 533; Child = 527; Teen = 499								

Note. ¶p < .10; * p < .05; ** p < .01; *** p < .001; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; r = correlated factors; bold font indicates best model by fit, factor discriminant validity, interpretability, consistency across sex and age sub-samples, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 7

Supervision/Involvement Scale: Initial measurement models

	Standardized factor loadings																	
	Monitoring						*Direct Supervision*						*Involvement*					
	CFA	*ESEM*					*CFA*	*ESEM*					*CFA*	*ESEM*				
Items	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*
1[A]		.17*	.18*	.09	-.01	.25		-.14*	-.28*	.00	-.01	-.20	.63*	**.67***	**.64***	**.65***	**.59***	**.71***
2[A]		.16*	.14*	.14*	-.02	.28		-.09*	-.28*	.09	.03	-.18	.70*	**.71***	**.75***	**.65***	**.67***	**.77***
3[B]	.85*	**.90***	**.90***	**.89***	**.90***	**.86***		.01	.04	-.06	.13*	.09		-.10*	-.02	-.04	-.01	.05
3a		-.06	-.05	-.11	.00	-.02	.61*	**.74***	**.60***	**.85***	**.69***	**.56***		-.20*	-.07	-.27*	-.06	-.13
4[B]	.74*	**.94***	**.93***	**.95***	**.89***	**.91***		-.13*	-.08*	-.17*	.05	.01		-.19*	-.08*	-.17*	-.11*	-.03
4a		.05	.03	.07	.02	.02	.85*	**.94***	**.86***	**1.01***	**.96***	**.69***		-.19*	-.10*	-.27*	-.02	.00
5[B]	.65*	**.71***	**.69***	**.71***	**.74***	**.58***		.08	.11	.12	-.07	.18		.01	.08	.02	.02	.26*
6[B]	.46*	**.60***	**.60***	**.59***	**.62***	**.41***		-.13*	-.10	-.16	-.22*	.12		.04	.15*	.04	-.04	.27*
7[B]	.80*	**.41***	.35*	**.51***	**.46***	.29*		**.40***	.39*	.49*	-.15*	**.59***		.21*	.28*	.16*	.40*	.20*
8[B]	.72*	**.60***	**.58***	**.61***	**.54***	**.54***		.09	.13	.09	-.33*	.30*		.23*	.32*	.21*	.26*	.27*
9[B]	.89*	**.62***	**.64***	**.60***	**.65***	**.60***		.30*	.30*	.35*	-.20*	.37*		.21*	.29*	.21*	.37*	.26*
10[B]	.72*	**.36***	**.42***	.30*	.13	**.49***		.30*	.28*	**.34***	-.13	.39*		.27*	.35*	.26*	**.44***	.20*
11		.26*	.30*	.19*	**.33***	.16	.58*	**.49***	**.46***	**.50***	.19*	**.64***		-.04	.03	-.02	.08	-.03
11a		-.15*	-.13	-.15	-.15	-.04	.46*	**.46***	**.48***	**.51***	.02	.27*		.09	.11	.01	**.18**	.09
12		.10*	.09	.06	.05	.17	.89*	**.76***	**.75***	**.71***	**.72***	**.75***		.08*	.05	.17*	.43*	.05
12a																		
13		-.06	-.13*	-.04	-.05	-.03	.76*	**.72***	**.77***	**.64***	**.59***	**.72***		.10*	.03	.21*	.41*	.07
13a																		
14		.25*	.14*	.34*	.19*	.19*		.39*	.41*	.34*	.03	**.53***	.72*	**.40***	**.44***	**.42***	**.58***	.36*
15		.20*	.18*	.15*	.08	.19*		.24*	.28*	.16*	.12	**.37***	.56*	**.37***	**.32***	**.49***	**.51***	.34*
16[A]		.05	.03	.04	.05	-.15		-.03	-.14*	-.04	-.29*	.14	.73*	**.76***	**.72***	**.81***	**.71***	**.77***
17[A]		-.08*	-.15*	-.05	.06	-.32*		.19*	.10	.18*	-.25*	.26*	.81*	**.77***	**.76***	**.78***	**.70***	**.83***
18[A]		.13*	.11*	.07	-.02	.07		-.04	-.09	-.13*	.02	.09	.71*	**.71***	**.74***	**.80***	**.72***	**.73***
19[A]		.02	-.07	.02	.01	-.14		.14*	.06	.08	-.05	.18	.83*	**.78***	**.79***	**.83***	**.81***	**.81***

20[A]		-.12*	-.25*	-.01	.00	-.24*		.41*	.35*	.40*	-.09	.43*	.72*	**.57***	**.58***	**.55***	**.61***	**.60***
21[A]		.02	.01	.02	-.10	.11		.27*	.20*	.29*	-.19*	.27*	.50*	**.36***	**.43***	**.34***	**.38***	**.39***
22[A]		.05	-.06	.09	-.04	-.06		.15*	.16*	-.01	.13*	.21	.81*	**.74***	**.79***	**.80***	**.84***	**.76***
23[A]		-.01	-.08	.00	-.06	-.15		.15*	.16*	-.01	-.03	.26*	.70*	**.65***	**.68***	**.71***	**.65***	**.69***
24		-.17*	-.21*	-..17*	-.17*	-.23*		**.27***	.25*	**.26***	.00	**.29***	.27*	.21*	**.26***	.15*	**.18***	.24*
25[A]		.18*	.14*	.14*	-.07	.26*		.09*	-.04	.17*	.11	.14	.77*	**.67***	**.77***	**.63***	**.72***	**.68***
30[A]		.00	-.09	.06	.01	-.10		.10*	.10	.03	.04	.11	.63*	**.60***	**.59***	**.61***	**.58***	**.64***
ω	.90	.89	.89	.89	.87	.88	.85	.85	.83	.87	.76	.79	.93	.92	.92	.92	.91	.93
α	.73	.73	.78	.72	.74	.73	.72	.72	.71	.73	.55	.69	.87	.87	.86	.89	.84	.89

1. You discuss with your child her/his plans for the coming day.[A]
2. You talk with your child about what s/he has actually done during the day.[A]
3. Your child has a set time to be home on school nights.[B]
3a. If there is a set time, what is it? ___ (RS)
4. Your child has a set time to be home on weekend nights.[B]
4a. If there is a set time, what is it? ___ (RS)
5. If your child did not come home by a time that was set, you would know.[B]
6. If you or another adult are not at home, your child leaves a note or calls to let you know where s/he is going.[B]
7. You know who your child's companions are when s/he is not at home.[B]
8. When you are not at home, your child knows how to get in touch with you.[B]
9. When your child is out, you know what time s/he will be home.[B]
10. It is important to you to know what your child is doing when s/he is outside of the home.[B]
11. Your child usually goes home afterschool.
11a. Is s/he usually supervised? (Y/N)
12. Your child is usually at home in the evenings.
13. Your child is usually at home on the weekends.
13a. Is s/he usually supervised? (Y/N)
14. You know many of your child's friends.
15. When you and your child are both at home, you know what s/he is doing.
16. Your child helps to plan family activities.[A]
17. Your child likes to get involved in family activities.[A]
18. You find time to listen to your child when s/he talks to you.[A]
19. You and your child do things together at home.[A]
20. Your child goes with members of the family to movies, sports events or other outings.[A]
21. Your child goes with members of the family to church, synagogue, or Sunday School.[A]
22. You often have a friendly chat with your child.[A]
23. Your child helps you.[A]
24. Your child prefers to be with her/his friends rather than with the family. (RS)
25. You talk with your child about how s/he is doing in school.[A]
30. In general, these activities are enjoyable?[A]

12a. Is s/he usually supervised? (Y/N)						
	Factor correlations					
	CFA	*ESEM*				
Factor pairs	*Full*	*Full*	*Boy*	*Girl*	*Child*	*Teen*
Monitoring and Direct Supervision	.36*	.16*	.06	.15*	.03	.29*
Monitoring and Involvement	.47*	.20*	.10*	.25*	.11	.24*
Direct Supervision and Involvement	.46*	.30*	.27*	.39*	-.07	.11
Sample Size: Full = 1,130; Boy = 495; Girl = 533; Child = 527; Teen = 499						

Note. * $p < .05$; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; [A] = original S/I Involvement item from Loeber and colleagues (1998); [B] = original S/I Supervision item from Loeber and colleagues (1998); bolded loadings = highest loading for each item within the full-sample or a sub-sample; boxes mark the factor that each item preferentially loads upon in ESEM; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006); (RS) = reverse-scored; (Y/N) = yes/no.

Table 8

Supervision/Involvement Scale best model comparisons: ESEM vs. CFA

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
ESEM Full	**1369(322)*****	**.916**	**.894**	**.054 [.051, .057]**	__	__	__	__
CFA Full	2289(374)***	.846	.833	.067 [.065, .070]	725(52)***	-0.07	-0.061	0.013
ESEM Boy	**729(322)*****	**.925**	**.906**	**.051 [.046, .055]**	__	__	__	__
CFA Boy	1191(374)***	.850	.837	.066 [.062, .071]	368(52)***	-0.075	-0.069	0.015
ESEM Girl	**815(322)*****	**.936**	**.920**	**.054 [.049, .058]**	__	__	__	__
CFA Girl	1384(374)***	.869	.858	.071 [.067, .075]	470(52)***	-0.067	-0.062	0.017
ESEM Child	**646(322)*****	**.908**	**.884**	**.044 [.039, .049]**	__	__	__	__
CFA Child	885(374)***	.855	.842	.051 [.047, .055]	230(52)***	-0.053	-0.042	0.007
ESEM Teen	**849(322)*****	**.934**	**.917**	**.057 [.053, .062]**	__	__	__	__
CFA Teen	1329(374)***	.880	.870	.072 [.067, .076]	413(52)***	-0.054	-0.047	0.015
Sample Size: Full = 1,130; Boy = 495; Girl = 533; Child = 527; Teen = 499								

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; bold font indicates best model by fit, factor discriminant validity, interpretability, consistency across sex and age sub-samples, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 9

Supervision/Involvement Scale: Measurement and structural invariance across sex and age

Model	*χ^2(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*	*$\Delta\chi^2$(df)*	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
Sex								
Configural	$1541(644)^{***}$	.931	.914	.052 [.049, .055]	__	__	__	__
Weak	$1408(722)^{***}$	.948	.941	.043 [.040, .046]	$107(78)^{**}$	0.017	0.027	-0.009
Strong	$1449(799)^{***}$	.950	.949	.040 [.037, .043]	100(77)	0.002	0.008	-0.003
Strict	$1438(828)^{***}$	.953	.954	.038 [.035, .041]	45(29)	0.003	0.005	-0.002
Factor Variance-Covariance	$1364(834)^{***}$	.960	.961	.035 [.032, .038]	$19(6)^{**}$	0.007	0.007	-0.003
Factor Means	**$1354(837)^{***}$**	**.961**	**.962**	**.035 [.031, .038]**	**6(3)**	**0.001**	**0.001**	**0.000**
Age								
Configural	$1424(644)^{***}$	.930	.911	.049 [.045, .052]	__	__	__	__
Weak	$1560(722)^{***}$	.924	.915	.048 [.044, .051]	$245(78)^{**}$	-0.006	0.004	-0.001
Strong	$1699(786)^{***}$	.918	.915	.048 [.044, .051]	$226(64)^{**}$	-0.006	.000	.000
Strict	$1709(815)^{***}$	.919	.920	.046 [.043, .049]	$81(29)^{**}$	0.001	+0.005	-0.002
Factor Variance-Covariance	**$1777(821)^{***}$**	**.914**	**.915**	**.048 [.045, .051]**	**$50(6)^{**}$**	**-0.006**	**-0.005**	**+0.002**
Factor Means	$2200(824)^{***}$	.876	.878	.057 [.054, .060]	$132(3)^{**}$	-0.038	-0.037	+0.009

Sample Size: Boy = 495; Girl = 533; Child = 527; Teen = 499

Note. ¶ p < .10; * p < .05; ** p < .01; *** p < .001; bold font indicates maximum level of invariance that cannot be rejected; a = item 13 thresholds and uniqueness non-invariant; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 10

Supervision/Involvement Scale: Final ESEM measurement models

	Standardized factor loadings					
	Monitoring		*Direct Supervision*		*Involvement*	
	All Invariant	*All invariant (-means)*	*All Invariant*	*All invariant (-means)*	*All Invariant*	*All invariant (-means)*
Items	*Boy Girl*	*Child Teen*	*Boy Girl*	*Child Teen*	*Boy Girl*	*Child Teen*
1	.14*	.13*	-.20*	-.13*	**.66***	**.63***
2	.15*	.14*	-.13*	-.10*	**.73***	**.70***
3	**.88***	**.87***	.04	.03	.00	-.03
3a	-.12*	-.08	**.73***	**.70***	-.10*	-.18*
4	**.94***	**.95***	-.07*	-.05*	-.10*	-.14*
4a	.00	-.05	**.95***	**.91***	-.08*	-.10*
5	**.69***	**.66***	.12*	.04	.10	.17*
6	**.59***	**.61***	-.12*	-.23*	.10*	.24*
7	.39*	**.42***	**.42***	.29*	.33*	.34*
8	**.58***	**.61***	.09	.07	.32*	.31*
9	**.60***	**.64***	.31*	.23*	.33*	.32*
10	.33*	**.42***	.28*	.27*	**.37***	.31*
11	.21*	.31*	**.48***	**.39***	.07	.07
11a	-.16*	-.14*	**.46***	**.32***	.12	.13
12	.04	.20*	**.69***	**.65***	.23*	.18*
12a						
13	-.11*	.03*	**.65***	**.61***	.23*	.20*
13a						
14	.22*	.26*	.35*	.30*	**.51***	**.49***
15	.14*	.20*	.19*	.21*	**.45***	**.41***
16	.02	-.02	-.12*	-.11*	**.78***	**.81***
17	-.12*	-.14*	.10*	.05	**.81***	**.87***
18	.07	.10*	-.13*	-.06	**.77***	**.74***
19	-.04	-.02	.03	.02	**.84***	**.83***
20	-.15*	-.12*	.32*	.25*	**.63***	**.67***
21	-.01	.05	.21*	.17*	**.42***	**.42***

22	-.02	.03	.05	.07*	**.81***	**.80***
23	-.07	-.03	.04	.05	**.70***	**.72***
24	-.20*	-.17*	.22*	.17*	**.23***	**.26***
25	.12*	.19*	.04	.07	**.73***	**.68***
30	-.04	-.02	.03	.03	**.62***	**.63***
ω	.89	.90	.84	.80	.93	.93
α	.74	.74	.72	.72	.87	.87

1. You discuss with your child her/his plans for the coming day.
2. You talk with your child about what s/he has actually done during the day.
3. Your child has a set time to be home on school nights.
3a. If there is a set time, what is it? ___ (RS)
4. Your child has a set time to be home on weekend nights.
4a. If there is a set time, what is it? ___ (RS)
5. If your child did not come home by a time that was set, you would know.
6. If you or another adult are not at home, your child leaves a note or calls to let you know where s/he is going.
7. You know who your child's companions are when s/he is not at home.
8. When you are not at home, your child knows how to get in touch with you.
9. When your child is out, you know what time s/he will be home.
10. It is important to you to know what your child is doing when s/he is outside of the home.
11. Your child usually goes home afterschool.
11a. Is s/he usually supervised? (Y/N)
12. Your child is usually at home in the evenings.
12a. Is s/he usually supervised? (Y/N)
13. Your child is usually at home on the weekends.
13a. Is s/he usually supervised? (Y/N)
14. You know many of your child's friends.
15. When you and your child are both at home, you know what s/he is doing.
16. Your child helps to plan family activities.
17. Your child likes to get involved in family activities.
18. You find time to listen to your child when s/he talks to you.
19. You and your child do things together at home.
20. Your child goes with members of the family to movies, sports events or other outings.
21. Your child goes with members of the family to church, synagogue, or Sunday School.
22. You often have a friendly chat with your child.
23. Your child helps you.
24. Your child prefers to be with her/his friends rather than with the family. (RS)
25. You talk with your child about how s/he is doing in school.
30. In general, these activities are enjoyable?

	Factor Correlations	
	All Invariant	*All Invariant (-means)*

	Boy	*Girl*	*Child*	*Teen*
Monitoring and Direct Supervision	.09*		.15*	
Monitoring and Involvement	.14*		.19*	
Direct Supervision and Involvement	.28*		.24*	
Sample Size: Boy = 495; Girl = 533; Child = 527; Teen = 499				

Note. *p < .05; (-means) = with the exception of factor means; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; boxes mark the factor that each item preferentially loads upon in ESEM; highest loading for each item is bolded for each multigroup model; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006); (RS) = reverse-scored; (Y/N) = yes/no.

Table 11

Emory Diagnostic Rating Scale – ODD: Factor analysis goodness-of-fit statistics

Model	χ^2*(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2$*(df)*	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
				ESEM				
3r Full	$75(7)^{***}$	.995	.981	.097 [.078, .117]	__	__	__	__
3r Boy	$55(7)^{***}$	.992	.966	.124 [.095, .156]	__	__	__	__
3r Girl	$52(7)^{***}$	.993	.973	.115 [.087, .145]	__	__	__	__
3r Child	$29(7)^{**}$	.996	.984	.081 [.052, .113]	__	__	__	__
3r Teen	$50(7)^{***}$	.994	.977	.115 [.086, .147]	__	__	__	__
2r Full	**$165(13)^{***}$**	**.989**	**.977**	**.106 [.092, .121]**	**$91(6)^{***}$**	**-0.006**	**-0.004**	**0.009**
2r Boy	**$109(13)^{***}$**	**.983**	**.964**	**.128 [.107, .151]**	**$57(6)^{***}$**	**-0.009**	**-0.002**	**0.004**
2r Girl	**$73(13)^{***}$**	**.991**	**.981**	**.097 [.076, .119]**	**$30(6)^{***}$**	**-0.002**	**0.008**	**-0.018**
2r Child	**$52(13)^{***}$**	**.993**	**.985**	**.079 [.057, .102]**	**$24(6)^{***}$**	**-0.003**	**0.001**	**-0.002**
2r Teen	**$85(13)^{***}$**	**.990**	**.979**	**.110 [.088, .133]**	**$40(6)^{***}$**	**-0.004**	**0.002**	**-0.005**
				CFA Modified Bifactor Model vs. Best ESEM Model				
MB Full	**$82(11)^{***}$**	**.995**	**.987**	**.079 [.063, .095]**	__	__	__	__
2r Full	$165(13)^{***}$	.989	.977	.106 [.092, .121]	$68(2)^{***}$	-0.006	-0.01	0.027
MB Boy	**$74(11)^{***}$**	**.989**	**.972**	**.113 [.090, .139]**	__	__	__	__
2r Boy	$109(13)^{***}$	.983	.964	.128 [.107, .151]	$38(2)^{***}$	-0.006	-0.008	0.015
MB Girl	**$48(11)^{***}$**	**.995**	**.986**	**.082 [.059, .107]**	__	__	__	__
2r Girl	$73(13)^{***}$	.991	.981	.097 [.076, .119]	$26(2)^{***}$	-0.004	-0.005	0.015
MB Child	**$38(11)^{***}$**	**.995**	**.988**	**.071 [.047, .097]**	__	__	__	__
2r Child	$52(13)^{***}$	.993	.985	.079 [.057, .102]	$17(4)^{**}$	-0.002	-0.003	0.008
MB Teen	**$70(11)^{***}$**	**.992**	**.980**	**.108 [.084, .132]**	__	__	__	__
2r Teen	$85(13)^{***}$	.990	.979	.110 [.088, .133]	$25(4)^{***}$	-0.002	-0.001	0.002

Sample Size: Full = 1,037 ; Boy = 447; Girl = 493; Child = 479; Teen = 459

Note. ¶p < .10; * p < .05; ** p < .01; *** p < .001; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; r = correlated factors; bold font indicates best model by fit, factor discriminant validity, interpretability, consistency across sex and age sub-samples, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 12

Emory Diagnostic Rating Scale - ODD: Initial measurement models

	Best CFA															*Best ESEM*	
	General					*OB*					*Irritability*					*ODD N/A*	*ODDB*
Items	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Boy*	*Girl*	*Child*	*Teen*	*Full*	*Full*
1	**-.72***	**-.78***	.05	-.58	**-.77***						.42*	.25*	**.84***	**.62**	.25*	.22*	**.67***
2	**-.98***	**-.98***	-.18	**-.80***	**-.96***	.19	-.05	**.98***	.57	-.11						.06	**.87***
3	**-.69***	**-.79***	.17	-.39	**-.84***	.44*	.28	**.79***	**.72***	.06						.26*	**.63***
4	-.58*	**-.67***	.38	-.25	**-.74***	**.60***	.43*	**.78***	**.84***	.31*						**.57***	.32*
5	-.59*	**-.71***	.40	-.23	**-.81***	**.64***	.47*	**.78***	**.80***	.38*						**.61***	.30*
6	-.58*	**-.72***	.40	-.25	**-.77***						**.71***	.61*	**.80***	**.84***	.64*	**.75***	.21*
7	-.64*	**-.75***	.36	-.26	**-.88***						**.68***	.49*	**.89***	**.90***	.32*	**.68***	.30*
8	-.56*	**-.70***	.48	-.16	**-.85***	**.71***	.52*	**.77***	**.85***	.38*						**.74***	.18*
ω	.94	.95	.79	.83	.96	.86	.69	.94	.93	.56	.85	.79	.92	.90	.75	.91	.86
α	.92	.92	.93	.92	.93	.89	.88	.88	.88	.89	.85	.84	.85	.82	.87	.90	.84

1. Loses Temper	5. Blames others for her/his mistakes or misbehavior
2. Argues with Adults	6. Is touchy or easily annoyed by others
3. Actively disobeys rules or refuses adults' requests (for example, refuses to do chores at home)	7. Is angry and resentful
4. Does things on purpose to annoy other people	8. Is spiteful or tries to get back at others

	CFA factor correlations					*ESEM factor correlations*	
Factor Pairs	*Full*	*Boy*	*Girl*	*Child*	*Teen*		*Full*
General and OB	.00	.00	.00	.00	.00	ODD N/A and ODDB	.72*
General and Irritability	.00	.00	.00	.00	.00		
OB and Irritability	.89*	.74*	.95*	.95*	.65*		

Sample Size: Full = 1,130; Boy = 495; Girl = 533; Child = 527; Teen = 499

Note. $^*p < .05$; Full = full sample; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; OB = Oppositional Behavior factor; ODD N/A = ODD Negative/Antagonistic factor; ODDB = ODD Behavior factor; highest loading for each item is bolded for each group; boxes mark the factor that each item preferentially loads upon in ESEM; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006).

Table 13

Emory Diagnostic Rating Scale - ODD: Measurement and structural invariance across sex and age

Model	*χ^2(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*	*$\Delta\chi^2$(df)*	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
Sex								
Configural[a]	103(23)*	.994	.985	.086 [.069, .103]	__	__	__	__
Weak[a]	102(36)*	.995	.992	.062 [.048, .077]	11(13)	0.001	0.007	-0.024
Strong[a]	122(57)*	.995	.995	.049 [.037, .061]	34(21)	0.000	0.003	-0.013
Strict[a]	141(64)*	.994	.995	.050 [.039, .062]	20(7)*	-0.001	.000	0.001
Factor Variance-Covariance[a]	102(68)*	.997	.998	.033 [.018, .045]	5(4)	0.003	0.003	-0.017
Factor Means[a]	**131(71)***	**.995**	**.996**	**.042 [.031, .054]**	**14(3)***	**-0.002**	**-0.002**	**0.009**
Age								
Configural[b]	87.274(23)*	.995	.988	.077 [.060, .095]	__	__	__	__
Weak[b]	132.307(36)*	.993	.989	.076 [.062, .090]	51(13)*	-0.002	0.001	-0.001
Strong[c]	152.490(58)*	.993	.993	.059 [.048, .070]	42(22)*	0.000	0.004	-0.017
Strict[c]	161.264(64)*	.993	.994	.057 [.046, .068]	17(6)*	0.000	0.001	-0.002
Factor Variance-Covariance[c]	119.135(68)*	.996	.997	.040 [.028, .052]	6(4)	0.003	0.003	-0.017
Factor Means[c]	**105.000(71)***	**.997**	**.998**	**.032 [.018, .044]**	**3(3)**	**0.001**	**0.001**	**-0.008**

Sample Size: Boy = 495; Girl = 533; Child = 527; Teen = 499

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; bold font indicates maximum level of invariance that cannot be rejected; a = residual variance for item 2 fixed to zero in female group due to non-significant negative residual variance; b = residual variance for item 6 fixed to zero in the adolescent offspring group due to non-significant negative residual variance; c = item 6 and item 2 residual variance fixed to zero in the adolescent; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* =

change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change).

Table 14

Emory Diagnostic Rating Scale - ODD: Final CFA measurement models

	Standardized factor loadings											
	General				*OB*				*Irritability*			
	All Invariant (-Item 2 δ)		*All Invariant (-Item 2 & 6 δ)*		*All Invariant (-Item 2 δ)*		*All Invariant (-Item 2 & 6 δ)*		*All Invariant (-Item 2 δ)*		*All Invariant (-Item 2 & 6 δ)*	
Items	*Boy*	*Girl*	*Child*	*Teen*	*Boy*	*Girl*	*Child*	*Teen*	*Boy*	*Girl*	*Child*	*Teen*
1	**.69***		**-.79***						.45*		.21*	
2	**.94***	**.96***	**-.97***	**-.99***	.27		-.08					
3	**.65***		**-.80***		.51*		.22*					
4	.51*		**-.72***		**.67***		.43*					
5	.52*		**-.74***		**.69***		.45*					
6	.53*		**-.69***	**-.74***					**.74***		.64*	.68*
7	.60*		**-.83***						**.73***		.41*	
8	.48*		**-.76***		**.75***		.48*					
ω	.93	.94	.96	.96	.88	.89	.71	.72	.86		.73	.77
α	.93	.93	.93	.93	.88	.88	.88	.88	.85		.85	.85

1. Loses Temper
2. Argues with Adults
3. Actively disobeys rules or refuses adults' requests (for example, refuses to do chores at home)
4. Does things on purpose to annoy other people
5. Blames others for her/his mistakes or misbehavior
6. Is touchy or easily annoyed by others
7. Is angry and resentful
8. Is spiteful or tries to get back at others

	CFA factor correlations			
Factor pairs	*Boy*	*Girl*	*Child*	*Teen*
General and OB	.00		.00	
General and Irritability	.00		.00	
OB and Irritability	.91*		.74*	

Sample Size: Boy = 495; Girl = 533; Child = 527; Teen = 499

Note. *$p < .05$; Boy = male sub-sample; Girl = female sub-sample; Child = child sub-sample; Teen = adolescent sub-sample; (-Item 2 δ) = with the exception of item 2 residual invariance; (-Item 2 & 6 δ) = with the exception of item 2 and 6 residual; highest loading for each item is bolded for each multigroup model; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates $\geq .90$ have been recommended (Lance et al., 2006).

Table 15

Emory Diagnostic Rating Scale - CD: Factor analysis goodness-of-fit statistics

Model	*χ^2(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*	*$\Delta\chi^2$(df)*	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
ESEM (14 items)								
3r	86(45)***	.994	.987	.029 [.020, .039]	—	—	—	—
2r	**119(57)*****	**.990**	**.984**	**.032 [.024, .041]**	**38(12)*****	**-.004**	**-.003**	**+.003**
ESEM (11 items)								
3r	19(18)	1.000	1.000	.006 [.000, .028]	—	—	—	—
2r	**43(27)***	**.998**	**.996**	**.024 [.008, .037]**	**22(9)****	**-0.002**	**-.004**	**+.018**
CFA (11 items)								
Modified Bifactor[a]	37(25)¶	.998	.997	.021 [.000, .035]	—	—	—	—
Bifactor[a]	62(26)***	.995	.990	.037 [.025, .048]	19(1)***	-0.003	-0.007	0.016
2r	**141(36)*****	**.987**	**.979**	**.053 [.044, .062]**	**70(10)*****	**-0.008**	**-0.011**	**0.016**
2o	1893(37)***	.761	.645	.220 [.211, .228]	448(1)***	-0.226	-0.334	+0.167
1 (vs. 2r)	139(37)***	.987	.980	.052 [.043, .061]	3(1)	0.000	0.001	-0.001
Best 11 Item Model Comparisons: ESEM vs. CFA								
2r ESEM	**43(27)***	**.998**	**.996**	**.024 [.008, .037]**	—	—	—	—
2r CFA	141(36)***	.987	.979	.053 [.044, .062]	86(9)***	-0.011	-0.017	0.029
Sample Size: 1,039								

Note. ¶ p < .10; * p < .05; ** p < .01; *** p < .001; r = correlated factors; bold font indicates best model by fit, factor discriminant validity, interpretability, reliability, and prior research; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (ΔTLI decrease > .10 = significant change); $\Delta RMSEA$ change in *RMSEA* relative to preceding model ($\Delta RMSEA$ increase > .015 = significant change).

Table 16

Emory Diagnostic Rating Scale - CD: Alternative measurement models

	Standardized factor loadings				
	One General	*Rule-Breaking*		*Aggression*	
	CFA	*CFA*	*ESEM*	*CFA*	*ESEM*
1 – Lies to get what he or she wants[a]	.68*	.71*	.39*		**.45***
2 – Lies to get out of trouble[a]	.68*	.71*	.29*		**.52***
3 – Lies to get out of responsibilities[a]	.68*	.71*	.37*		**.46***
4 – Lies to get others in trouble[a]	.78*	.81*	.09		**.73***
5 – Bullies or threatens	.85*		.16*	.86*	**.77***
6 – Starts physical fights with people at home[b]	.83*		-.08	.85*	**.94***
7 - Starts fights with people who do not live at home[b]	.68*		.24*	.69*	**.54***
8 - Skipped school or work	.55*	.56*	**.96***		-.17*
9 - Ran away from home overnight					
10 - Stole items worth more than $20, but without force or threat towards another person (e.g., shoplifting, forgery)					
11 - Destroyed others' property on purpose (other than by setting fire)	.66*	.67*	.33*		**.46***
12 - Set fires wanting to cause serious damage					
13 - Broke into someone else's house, building, or car					
14 - Was physically cruel to animals					
15 - Was physically cruel to people	.77*		.30*	.78*	**.60***
16 - Stole things from another person by using force or threat (e.g., mugging, purse snatching, extortion)					
17 - Has used a weapon that could seriously harm others (e.g., brick, bat, knife, gun)					
18 - Forced someone into sexual activity with her/him					
19 - Stayed out late against parents' wishes	.60*	.61*	**.77***		.03
ω	.92	.86	.84	.87	.87
α	.82	.79	.60	.64	.74

	Factor correlations	
	CFA	ESEM
Rule-Breaking with Aggression	.94*	.42*

	ESEM	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*	*ESEM*	*CFA*
	Item residual correlations (2-factor oblique models)											
	Item 1[a]		Item 2[a]		Item 3[a]		Item 4[a]		Item 6[b]		Item 7[b]	
Item 1[a]	1.0	1.0										
Item 2[a]	.80*	.74*	1.0	1.0								
Item 3[a]	.59*	.57*	.68*	.66*	1.0	1.0						
Item 4[a]	.34*	.19	.43*	.39*	.38*	.28*	1.0	1.0				
Item 6[b]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0		
Item 7[b]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-.10	-.04	1.0	1.0
Sample Size: 1,039												

Note. $^{*}p < .05$; a = item residual variance allowed to correlate among the items making up the CD lying symptom; b = item residual variance allowed to correlate among the items making up the CD initiating physical fights symptom; highest loading for each item in each version of the scale is bolded; boxes mark the factor that each item preferentially loads upon in ESEM; ω = McDonald's (1970) scale score reliability coefficient; α = Cronbach's alpha (1951) scale score reliability estimate; reliability estimates ≥ .70 are traditionally considered acceptable and estimates ≥ .80 are traditionally considered good (Lance, Butts, & Michels, 2006), but in applied settings where decisions based on test scores are of high consequence, estimates ≥ .90 have been recommended (Lance et al., 2006).

Table 17

Emory Diagnostic Rating Scale - CD: MIMIC model with sex and age main effects

Model	$\chi^2(df)$	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2(df)$	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
Null	$185(49)^{***}$	.982	.972	.054 [.046, .063]	—	—	—	—
Threshold Invariance	**$78(45)^{**}$**	**.996**	**.993**	**.028 [.017, .038]**	**$50(4)^{***}$**	**0.014**	**0.021**	**-0.026**
Partial Threshold Invariance	$75(44)^{**}$	.996	.993	.027 [.016, .038]	$3(1)^{¶}$	.000	.000	-.001

Mean differences in Aggression and Rule-Breaking: Threshold Invariance MIMIC Model

	MIMIC Covariates β (s.e.)	
	Sex	Age
Rule-Breaking	-.02 (.06)	$.40^{***}$(.06)
Aggression	$-.19^{***}$(.05)	$-.33^{***}$ (.05)

Sample Size: 940

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; bold font indicates the best-fitting model based on incremental improvement in model fit; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 = significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change); β = path coefficient; (s.e.) = standard error.

Table 18

SEM of ODD, broad parenting dimensions, and specific parenting practices: Goodness-of-fit statistics

Model	χ^2*(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*	$\Delta\chi^2$*(df)*	*ΔCFI*	*ΔTLI*	*ΔRMSEA*
				Nurturance and Restrictiveness				
Baseline across sex	4246(2963)***	.922	.924	.029 [.027, .031]	__	__	__	__
β Invariance across sex	**4172(2969)*** **	**.927**	**.929**	**.028 [.026, .030]**	**3(6)**	**+.005**	**+.005**	**-.001**
Baseline across age	4234(2958)***	.920	.923	.029 [.027, .031]	__	__	__	__
β Invariance across age	**4192(2964)*** **	**.923**	**.926**	**.028 [.026, .030]**	**10(6)**	**+.003**	**+.003**	**-.001**
				Sample Size: Boy = 496; Girl = 535; Child = 528; Teen = 501				
				Monitoring, Direct Supervision, and Involvement				
Baseline across sex	2086(1376)***	.965	.966	.032 [.029, .034]	__	__	__	__
β Invariance across sex	**1972(1385)*** **	**.971**	**.972**	**.029 [.026, .031]**	**5(9)**	**+.006**	**+.006**	**-0.003**
Baseline across age	2428(1360)***	.945	.946	.039 [.037, .042]	__	__	__	__
β Invariance across age	**2436(1369)*** **	**.945**	**.946**	**.039 [.036, .041]**	**36(9)*** **	**.000**	**.000**	**.000**
				Sample Size: Boy = 497; Girl = 535; Child = 528; Teen = 502				
				Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement				
Baseline across sex	8731(6917)***	.922	.923	.023 [.021, .024]	__	__	__	__
Factor r invariance across sex	8636(6923)***	.926	.927	.022 [.020, .023]	8(6)	+.004	+.004	-.001
β Invariance across sex	**8554(6938)*** **	**.930**	**.931**	**.021 [.020, .023]**	**10(15)**	**+.004**	**+.004**	**-.001**
Baseline across age	8837(6896)***	.911	.912	.023 [.022, .025]	__	__	__	__
Factor r invariance across age	8791(6902)***	.913	.915	.023 [.022, .024]	16(6)*	+.002	+.003	.000
β Invariance across age	**8824(6917)*** **	**.913**	**.914**	**.023 [.022, .025]**	**40(15)*** **	**.000**	**-.001**	**.000**
				Sample Size: Boy = 497; Girl = 535; Child = 528; Teen = 502				

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; r = correlation; bold font indicates maximum level of invariance that cannot be rejected; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; $\Delta\chi^2$ = change in χ^2 relative to the preceding model; *ΔCFI* = change in *CFI* relative to preceding model (*ΔCFI* decrease > .10 = significant change); *ΔTLI* = change in *TLI* relative to preceding model (*ΔTLI* decrease > .10 =

significant change); *ΔRMSEA* change in *RMSEA* relative to preceding model (*ΔRMSEA* increase > .015 = significant change); β = path coefficient; (s.e.) = standard error.

Table 19

SEM ODD, broad parenting dimensions, and specific parenting practices: Prediction invariance

	Nurturance and Restrictiveness β (s.e.)			*Monitoring, Direct Supervision, and Involvement β (s.e.)*			
	Nurturance	Restrictiveness	%Variance explained (s.e.)	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
	Boy / Girl	Boy / Girl	Boy / Girl	Boy / Girl	Boy / Girl	Boy / Girl	Boy / Girl
General	-.12**(.04)	.04(.05)	2*(1)	.07(.05)	.08¶(.05)	-.29***(.04)	8**(2)
Irritability	-.22***(.05)	.17**(.05)	7**(3)	.06(.05)	.09¶(.05)	-.21***(.05)	4*(2)
OB	-.23***(.05)	.22***(.05)	10**(3)	.00(.05)	.22***(.05)	-.30***(.05)	10**(3)
	Sample Size: Boy = 496; Girl = 535			*Sample Size:* Boy = 497; Girl = 535			
	Nurturance	Restrictiveness	%Variance explained (s.e.)	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
	Child / Teen	Child / Teen	Child / Teen	Child / Teen	Child / Teen	Child / Teen	Child / Teen
General	-.22***(.04)	.13**(.04)	7**(2)	.07(.04)	.11**(.04)	-.40***(.04)	15***(3)
Irritability	-.07(.05)	.08(.05)	1(1)	.05(.05)	-.07(.05)	.05(.05)	.8(.8)
OB	-.14*(.05)	.18**(.06)	5(3)	-.02(.06)	.17**(.05)	-.20**(.06)	5(3)
	Sample Size: Child = 528; Teen = 501			*Sample Size:* Child = 528; Teen = 502			

	Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement β (s.e.)						
	Nurturance	Restrictiveness		Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
	Boy / Girl	Boy / Girl		Boy / Girl	Boy / Girl	Boy / Girl	Boy / Girl
General	.14¶(.08)	.03(.05)		.05(.05)	.07(.05)	-.38***(.08)	8**(3)
Irritability	-.16¶(.09)	.16**(.05)		.07(.05)	.06(.06)	-.10(.09)	8**(3)
OB	-.03(.09)	.20***(.05)		-.01(.05)	.17**(.06)	-.28**(.09)	14***(4)
	Nurturance	Restrictiveness		Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
	Child / Teen	Child / Teen		Child / Teen	Child / Teen	Child / Teen	Child / Teen
General	.04(.10)	.11**(.04)		.05(.05)	.10*(.04)	-.47***(.07)	16***(3)
Irritability	-.26**(.09)	.07(.06)		.07(.05)	-.09¶(.05)	.25**(.09)	4*(2)
OB	-.04(.10)	.18**(.06)		-.04(.06)	.15**(.05)	-.16(.11)	8*(3)
	Sample Size: Boy = 497; Girl = 535; Child = 528; Teen = 502						

Predictors correlation matrix (sex below line, age above)

Outcomes correlation matrix

	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement
Nurturance	1.0	.02^{A}	.25***	.15***	.76***
Restrictiveness	.03^{A}	1.0	.12**	.10*	.02
Monitoring	.21***	.10*	1.0	.15^{A}	.19^{A}
Direct Supervision	.13**	.16***	.09^{A}	1.0	.24^{A}
Involvement	.77***	.04	.14^{A}	.28^{A}	1.0

(sex below, age above)

	General	Irritability	OB
General	1.0	0.0	0.0
Irritability	0.0	1.0	.74^{A}
OB	0.0	.91^{A}	1.0

Note. ¶ p < .10; * p < .05; ** p < .01; *** p < .001; β = path coefficient; (s.e.) = standard error; A = factor correlation fixed to its value in original measurement and structural invariance model.

Table 20

SEM of CD, age, sex, broad parenting dimensions, and specific parenting practices

Model	*N*	χ^2*(df)*	*CFI*	*TLI*	*RMSEA [90%CI]*
Sex and Age	940	97(48)***	.994	.990	.033 [.024, .043]
Sex, Age, Nurturance, and Restrictiveness	1,029	2922(1590)***	.893	.885	.029 [.027, .030]
Sex, Age, Monitoring, Direct Supervision, and Involvement	1,030	2064(750)***	.913	.900	.041 [.039, .043]
Sex, Age, Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement	1,030	5717(3640)***	.898	.893	.024 [.022, .025]

Sex and Age

	Covariates β *(s.e.)*		**Predictors** β *(s.e.)*					
Outcomes	Sex	Age	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
Rule-Breaking	-.02(.06)	.40***(.06)						16***(5)
Aggression	-.19***(.05)	-.33***(.05)						14***(4)

Sex, Age, Nurturance, and Restrictiveness

Outcomes	Sex	Age	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
Rule-Breaking	-.01(.06)	.42***(.06)	-.32***(.05)	.06(.06)				28***(5)
Aggression	-.19***(.05)	-.31***(.06)	-.20***(.05)	.17**(.05)				20***(4)
Non-Invariant T								
Item 13 M-CRPR		-.28***(.04)						

Sex, Age, Monitoring, Direct Supervision, and Involvement

Outcomes	Sex	Age	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)

Outcomes	Sex	Age	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
Rule-Breaking	-.02(.06)	.46***(.05)			.02(.06)	-.33***(.06)	-.32***(.05)	48***(6)
Aggression	-.18***(.05)	-.26***(.06)			.10(.05)	.06(.06)	-.38***(.04)	23***(4)

Sex, Age, Nurturance, Restrictiveness, Monitoring, Direct Supervision, and Involvement

Outcomes	Sex	Age	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement	%Variance explained (s.e.)
Rule-Breaking	-.02(.06)	.46***(.05)	.02(.08)	.06(.06)	-.11*(.05)	-.25***(.06)	-.31***(.08)	44***(6)
Aggression	-.17***(.05)	-.23***(.06)	.06(.07)	.17**(.05)	.05(.05)	.06(.06)	-.43***(.07)	25***(4)
Non-Invariant T								
Item 13 M-CRPR		-.28***(.04)						

Predictors correlation matrix

	Nurturance	Restrictiveness	Monitoring	Direct Supervision	Involvement
Nurturance	1.0				
Restrictiveness	.01	1.0			
Monitoring	.29***	.12*	1.0		
Direct Supervision	.27***	.11*	.16	1.0	
Involvement	.76***	.01	.20	.30	1.0

Outcomes correlation matrix

	Rule-Breaking	Aggression
Rule-Breaking	1.0	
Aggression	.42	1.0

Note. ¶ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; χ^2 = robust weighted least squares chi-square; *CFI* = comparative fit index; *TLI* = Tucker-Lewis index; *RMSEA [90%CI]* = 90% confidence interval for the root mean square error of approximation point estimate; β = path coefficient; (s.e.) = standard error; T = item threshold.