Published in final edited form as:

Assessment. 2017 July ; 24(5): 646–659. doi:10.1177/1073191115620839.

Alabama Parenting Questionnaire-9: Longitudinal Measurement Invariance across Parents and Youth during the Transition to High School

Thomas J. Gross,

Tennessee State University, 313 Clay Hall, Nashville, TN 37209. University of Nebraska-Lincoln, 213 Barkley Memorial Center, Lincoln, NE 68583

Charles B. Fleming,

Social Development Research Group, University of Washington, Seattle, UW Box #358734, 9725 Third Avenue NE, Suite #401, Seattle, WA 98115

W. Alex Mason, and

National Research Institute for Child and Family Studies, Boys Town, NE, 14100 Crawford Street, Boys Town, NE 68010

Kevin P. Haggerty

Social Development Research Group, University of Washington, Seattle, UW Box #358734, 9725 Third Avenue NE, Suite #401, Seattle, WA 98115

Abstract

The Alabama Parenting Questionnaire nine-item short form (APQ-9) is an often used assessment of parenting in research and applied settings. It uses parent and youth ratings for three scales: Positive Parenting, Inconsistent Discipline, and Poor Supervision. The purpose of this study is to examine the longitudinal invariance of the APQ-9 for both parents and youth, and the multi-group invariance between parents and youth during the transition from middle school to high school. Parent and youth longitudinal configural, metric and scalar invariance for the APQ-9 were supported when tested separately. However, the multi-group invariance tests indicated that scalar invariance was not achieved between parent and youth ratings. Essentially, parent and youth mean scores for Positive Parenting, Inconsistent Discipline, and Poor Supervision can be independently compared across the transition from middle school to high school. However, comparing parent and youth scores across the APQ-9 scales may not be meaningful.

Keywords

Parenting assessment; longitudinal measurement invariance; multi-group measurement invariance; early adolescence; Alabama Parenting Questionnaire

Correspondence and inquiries may be sent to Thomas J. Gross, Tennessee State University, 313 Clay Hall, Nashville, TN 37209, tjgross44@gmail.com.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies or the National Institutes of Health or the U.S. Department of Education.

Parenting practices are frequently linked to youth behaviors, such as delinquency, social skills, and school performance, across all periods of child and adolescent development (Barry, Lochman, Fite, Wells, & Colder, 2012; Hoeve et al. 2009; Tyler, Johnson, & Brownridge, 2008). Researchers and clinicians, alike, typically assess parenting practices to intervene on parenting skills in order to provide comprehensive care for youth. It is common to use both parent and child ratings, but the correlations between the same survey questions is usually low (i.e., $r < .3$; Fleming, Mason, Thompson, Haggerty, & Gross, 2015; Jacob & Windle, 1999; Pasch, Stigler, Perry, & Komro, 2010).

Low convergence between parents and children on parenting assessments may be a matter of divergent perceptions between parents and children. Items related to parent behavior and family processes, such as family rules or consistency of consequences, may be better reported by parents (De Los Reyes, Goodman, Kliewer, & Reid-Quinones, 2010). Items related to child behavior driven processes, such as parental monitoring or supervision practices, may be more likely to capture what children are likely to disclose to their parents about whereabouts and peer groups (Kerr & Stattin, 2000; Kerr, Stattin, & Burk, 2010). Even for well-established parenting constructs, it is likely that parents and youth are reporting information regarding the same parenting practices, but parents and youth may respond to items with different degrees of understanding during adolescence (Janssens et al., 2015).

One instrument that was designed to assess parenting constructs across parents and children is the Alabama Parenting Questionnaire (APQ; Shelton, Frick, & Wootton, 1996). The APQ was designed to assess the parenting practices most related to disruptive behaviors in children (Shelton et al., 1996), and is frequently used in research and applied settings. The full form of the APQ consists of 42 items that measure (a) parental involvement (10 items), (b) positive parenting (6 items), (c) inconsistent discipline (6 items), (d) poor monitoring/supervision (9 items), and (e) corporal punishment (3 items). The APQ may be rated by parents and children aged 6 to 17 years (Frick, Christian, & Wootton, 1999; Shelton et al., 1996). An initial development study found that reliabilities for the parental involvement, positive parenting, and poor monitoring/supervision scales were acceptable for both parent and child raters ($\alpha = .67$ to $.80$; Shelton et al., 1996).

The 42-item format of the APQ makes it cumbersome to complete multiple times as part of a clinical or research assessment battery. Short forms that use a minimal number of items while achieving adequate internal consistency are often preferred for repeated assessment in order to prevent respondent fatigue and allow for the examination of other areas of interest (Boyle, Jadad, & Allnut, 1999; Elgar, Waschbusch, McGrath, Stewart, & Curtis, 2004). A 9-item short form was developed based on a three factor model and was named the APQ-9 (Elgar, Waschbusch, Dadds, & Sigvaldason, 2007). The three factors were extracted using an exploratory factor analysis that included all APQ items, and then the three items with the highest loadings for each factor were selected for the APQ-9 (Elgar et al., 2007). The APQ-9 has three three-item scales, which correlated highly their respective full-form scale: (a) Positive Parenting, $r = .89$; (b) Inconsistent Discipline, $r = .90$; and (c) Poor Supervision, $r = .76$. The initial development and validation study for the APQ-9 was conducted with parents only, and it was found that internal consistency varied across each of the APQ-9

scales for samples of parents with children across different age ranges (child age 4 to 9 years, across scales mean $\alpha = .44$; child age 5 to 12 years, range $\alpha = .59$ to $.84$; child age 5 to 18 years, range $\alpha = .57$ to $.61$; Elgar et al., 2007). However, a confirmatory factor analysis of the APQ-9 yielded good model fit for mothers (CFI = .99, NFI = .98) and fathers (CFI = .99, NFI = .98; Elgar et al., 2007). In fact, other investigations of the APQ have supported a three factor model that includes: (a) positive parenting, (b) inconsistent discipline, and (c) poor monitoring/supervision (Hawes & Dadds, 2006; Hinshaw et al., 2000; Molinuevo, Pardo, & Torrubia, 2011).

The intent of the APQ is to have meaningful parent and youth reports of parenting practices and to capture potential changes over time (Frick et al., 1999; Shelton et al., 1996), but the psychometrics of the APQ have had limited evaluations for youth and longitudinal evaluations of the instrument are generally scarce. The psychometrics for the APQ-9 have not been examined for youth. It is important to establish that rating scales may be shortened and continue to have adequate measurement properties, while sufficiently assessing the same factors across raters and over time (e.g., Gresham et al., 2010; Volpe & Gadow, 2010). This would allow for the use of the APQ-9 to provide a meaningful assessment of changes in parenting, and to understand the degree of correspondence across youth and parent reports. The APQ and APQ-9 have been employed in this manner in research studies, but without a comprehensive psychometric examination of their utility as a multi-informant longitudinal parenting assessment.

The APQ scale score correlations have been examined between parent and child ratings in some studies, but it is still uncertain if the parenting constructs are rated in the same way. In comparisons of youth and parent report, scale score correlations were modestly related between youth and parents, and scale means were found to be significantly different between youth and parents (Molinuevo et al., 2011; Scott Briskman, & Dadds, 2011). Further, there is some indication that the APQ items may relate to the scale constructs differently for youth (Essau, Sasagawa, & Frick, 2006). Studies that examined APQ ratings across youth age groups found trends across middle childhood (9 to 12 years) and adolescence (13 to 17 years) that indicated parent ratings had consistent or increasing internal consistency across these age groups, whereas youth ratings fluctuated (e.g., Poor Supervision internal consistency coefficients decreased from $\alpha = .72$ to $.43$; Frick et al., 1999). However, these age groups were studied in a cross-sectional manner, which leaves little information regarding if ratings change over time, and if so, in what ways. These findings suggest that thorough examinations should be conducted to establish not only if youth and parents scores correspond, but if they relate to the same underlying constructs within and across time.

Longitudinal and Multi-group Measurement Invariance

The ability of a scale to assess the same construct in a replicable manner is referred to as measurement invariance (Dimitrov, 2010). It is often assumed that assessments measure the same construct over time, which is known as *longitudinal invariance*, and across groups, which is known as *multi-group invariance*. Invariance is demonstrated through the confirmation of items (a) loading on their presumed constructs, *configural invariance*, (b) contributing similarly to the construct, *metric invariance*, and (c) having similar mean

structures, *scalar invariance*, both over time and between groups (Milfont & Fischer, 2010; van de Schoot, Lugtig, & Hox, 2012) in longitudinal multi-informant studies. Establishing configural invariance allows assessment users to use similar names for the scales because there is evidence that the same items group together across time and groups (Dimitrov, 2010). Establishing metric invariance provides assurance to assessment users that the strength of association of the items to the factor and associations of the items to each other are similar. This is important because it substantiates the generalization of assessment factor(s) across time and groups (Dimitrov, 2010). Establishing scalar invariance ensures that the relative means, or frequency distribution of items within constructs, follow a similar pattern. This provides information to inform assessment users that scale means are comparable across time and groups (Dimitrov, 2010).

Presumptions of longitudinal invariance ought to be tested because both parents and children undergo expected changes through the typical course of child development as roles and responsibilities change (Granic & Patterson, 2006). Separate tests of parent and youth longitudinal invariance would allow for conclusions regarding the consistency of each rater's perception of parenting from the beginning to the end of the transition to high school, via the APQ-9. Moreover, groups with expected differences, such as parents and children, may have differential responses to items due to distinctly different perceptions of how items and constructs are operationalized (Vandenberg & Lance, 2000). Understanding the degree to which the APQ-9 scores reflect convergence or divergence across parents and youth during this transitional period to high school may be useful to researchers and practitioners. Regardless of invariance or non-invariance, researchers and practitioners may make more informed comparisons and interpretations of parenting practices assessed through the APQ-9.

The Current Study

Multi-rater parenting measures that demonstrate invariance across developmental periods improve researchers' confidence in parent and child score comparisons over time. Significantly, the longitudinal investigation of parenting assessments could improve the understanding of parenting processes across child developmental periods. This may improve theory by confirming prior assumptions for different time periods as well as addressing gaps in the literature specifically regarding the middle school to high school transition. Additionally, this may help intervention impact assessment by improving our understanding of youth and parent mean differences across the transition from middle school to high school. The purpose of this study is to examine the longitudinal invariance of the APQ-9 for both parents and youth, and the multi-group invariance between parents and youth during the transition from middle school to high school.

Method

Participants

Participants included 321 low-income families enrolled in a randomized controlled trial of a parenting program, Common Sense Parenting (Burke, Herron, & Barnes, 2006). Each family included a target parent and target eighth grader, who attended one of five low-performing

middle schools in the urban Pacific Northwest. Just over 70% of all the students at all five schools were receiving free or reduced-price school lunch. Data were collected from the participants at three time points: baseline (middle of eighth grade), post-test (end of eighth grade), and follow-up (end of ninth grade).

Parent demographic data included age, parent relationship to the child, race, level of education, house hold income, and employment status. Race and ethnicity were viewed as distinct constructs; therefore, identification with Hispanic ethnicity was collected separate from race. Table 1 contains the parent demographic data at baseline and at follow-up. Youth demographic data included age, gender, race, and identification with Hispanic ethnicity. Table 2 contains the youth demographic data at baseline and at follow-up.

Alabama Parenting Questionnaire-9

The Positive Parenting, Inconsistent Discipline, and Poor Supervision scales from the Alabama Parenting Questionnaire-9 (APQ-9; Elgar et al., 2007) were used, and all scales contained three items rated on a 5-point scale (1 = *Never*; 2 = *Almost never*; 3 = *Sometimes*; 4 = *Often*; 5 = *Always*). The Positive Parenting scale was used to assess how often the parent provided praise and recognition for prosocial behavior. The Inconsistent Discipline scale assesses the frequency with which parents fail to follow through with consequences for misbehavior. The Poor Supervision scale assesses parental knowledge of child behaviors. All the same items from the scales were asked to the youth with minor changes in wording to reflect they were to report about their parents. Youth were given the set questions at each data collection period with instructions to rate their parents. Table 3 provides the wording for all items for parent and youth, grouped by subscale.

Procedure

Potential participants were informed of the project by research staff who presented the study during core classes and distributed permission-to-contact forms for the students to take home to their parents. Schools aided the recruitment effort by disseminating notices of the study (e.g., emails, automated phone reminders). Schools also mailed a copy of the permission-to-contact forms directly to families who had not responded to initial recruitment efforts.

Families were enrolled in the project in two cohorts, where the first cohort was recruited from three schools and the second cohort was recruited from five schools, of which three were the same as with the first data collection. The total population of eighth-grade students for the first and second cohort consisted of 1,646 students. Permission slips with agreement to release contact information, were returned by 658 families. Of these 658 families, 122 families in the 2010–2011 school year and 199 families in 2011–2012 were contacted, determined eligible, and chose to participate in the project (see Mason et al., 2014). All procedures were reviewed and approved by the *University* and *Agency* Institutional Review Boards as well as the participating school district.

Parents and students completed interviews that included baseline surveys when they enrolled in the parenting program. They completed post-test surveys approximately six months after the baseline surveys, and follow-up surveys one year after the post-test surveys. The surveys corresponded to middle of eighth grade, end of eighth grade, and end of ninth grade,

respectively. The APQ-9 was included within each participant's survey, and every attempt was made to ensure that the same parent identified as the primary caregiver completed the survey as the primary caregiver each time, although there were a few exceptions ($n = 32$; e.g., when a primary caregiver developed a debilitating illness and could no longer participate in the study. Enrollment and baseline interviews began in November/December and were completed before April. Post-test and follow-up interviews began in May/June and were completed before September. The intervention programs were delivered in various community settings (e.g., schools, churches), whereas the parent and child surveys were self-administered on laptop computers in families' homes, with a data collection staff member present to provide assistance during all data collection periods. The baseline, post-test, and follow-up data were used in this study. Overall, 286 (89%) of the families enrolled in the study completed all APQ-9 survey items at baseline, post-test, and follow-up. Follow-up completion rates did not differ by race, ethnicity, or whether families received food stamps. The completion rate for families of boys (97%) was significantly higher than that for families of girls (92%; $\chi^2 = 6.22, p = .049$).

Analyses

The participants were assigned to one of three conditions in a randomized controlled trial of the Common Sense Parenting (CSP) program. Families were assigned to a minimal contact control condition ($n = 108$), a standard six-session CSP condition ($n = 118$), and an eight-session CSP-*Plus* condition ($n = 95$). It was expected that parents receiving the CSP and CSP-*Plus* interventions would demonstrate improvements in parenting measured by the APQ-9 when compared to control condition participants; however, no such improvements have been found (Mason et al., 2015). A full discussion of intervention-related results goes beyond the scope of this paper and can be found elsewhere (e.g., Mason et al., 2015), although we note that the APQ-9 may have been insensitive to detecting change or the interventions may not have impacted the relatively established parenting behaviors of families with adolescent-age children. A repeated measures multivariate analysis of variance (MANOVA) indicated that the differences between groups on the parent APQ-9 were non-significant from baseline through follow-up, $F(12, 562) = 1.15; p = .315$; Wilk's $\Lambda = .95$. Similarly, a repeated measures MANOVA indicated that the differences between groups on the youth APQ-9 were non-significant from baseline through follow-up, $F(12, 566) = 1.17; p = .298$; Wilk's $\Lambda = .95$. Therefore, analyses were based on the total sample pooled across experimental conditions. During the initial analyses, each scale scores was calculated at each time point and used to examine item and scale means and correlations across reporters.

Differences in APQ-9 scores by the five middle schools from which students were recruited into the project and there was little evidence of level differences on either parent- or child-report APQ-9 scales. A repeated measures MANOVA indicated that the differences between schools on the parent APQ-9 were non-significant from baseline through follow-up, $F(24, 946.62) = 0.63; p = .912$; Wilk's $\Lambda = .95$, as were differences between schools on the youth APQ-9, $F(24, 1930.40) = 0.89; p = .614$; Wilk's $\Lambda = .96$. Based on the lack of significant differences between schools and the absence of any reason to expect relationships among APQ-9 items to differ by school, we did not include school in our primary analysis.

MANOVAs were used to examine if there were differences between parent- and youth-rated scores for those who completed all APQ-9 subscales at baseline and follow-up and those who only completed all APQ-9 subscales at baseline. The differences between parents who completed the survey at baseline and follow-up ($n = 302$) and those who did not complete at follow-up ($n = 16$) were non-significant, $F(3, 314) = 0.42$; $p = .763$; Wilk's $\Lambda = 1.00$. Similarly, youth who completed the surveys at baseline and follow-up ($n = 305$) and those who did not ($n = 15$) did not differ significantly, $F(3, 316) = 0.12$; $p = .947$; Wilk's $\Lambda = 1.00$.

There was a small amount of missing data at the individual item level (4%) and from the scale scores (4%), which is unlikely to contribute to bias (Graham, 2009). Still, primary analyses were performed in Mplus 7.11 (Muthén & Muthén, 1998–2013) using the full-information maximum likelihood estimator (FIML), which allows for inclusion of cases with partially missing data (Schafer & Graham, 2000). All item values were acceptable for skewness ($< |3|$) and kurtosis ($< |8|$; Kline, 2010) across all time points. The APQ-9 parent and youth forms were examined separately for longitudinal invariance across baseline, post-test, and follow-up with a series of sequential analyses to test the configural, metric, and scalar invariance assumptions within a confirmatory factor analysis (CFA) framework. In a longitudinal invariance comparison between groups, the first step is to test whether the proposed factor structure fits the data from each group (Little, 2013). Following the separate group analyses, multi-group longitudinal invariance across the three time points was examined using the same sequential testing to determine if invariance held across parents and youth, over time.

The effects coding approach was used for scaling the item factor loadings and item mean intercepts. In this approach constraints are used to set the factor loadings of each construct to average 1.0 and the corresponding item mean intercepts are constrained to sum to 0 (Little, 2013). The resulting variance of the latent variables reflects the mean of the corresponding items' variances explained by the construct. Further, this allows the latent variable means to reflect the average of the corresponding items' means, which scales the latent variable on the same metric as the manifest items (Little, 2013). The effects coding approach is advantageous for longitudinal and multi-group invariance testing because the unstandardized parameter estimates may be interpreted in a manner that is directly related to the scale of item measurement, rather than the arbitrary metric created when the marker method (setting one factor loading to 1.0) is used (Brown, 2015).

Within the configural invariance tests of the CFA model, the factor loadings or mean intercepts were allowed to be freely estimated over time for longitudinal invariance, and between parents and youth for multi-group invariance. The metric invariance tests constrained each item's unstandardized factor loading parameter estimate to equality over time for longitudinal invariance, and between parents and youth for multi-group invariance. The scalar invariance tests constrained each item's unstandardized factor loading parameter estimate and item mean intercepts to equality over time for longitudinal invariance, and between parents and youth for multi-group invariance.

The indicators used to assess goodness-of-fit were the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Tucker & Lewis, 1973), and Gamma-hat (Fan & Sivo, 2007), which range from 0 to 1.00. Gamma-hat was selected as a third fit index because it is sensitive to model misspecification, sensitizes model fit to the number of variables used in the model, and is less influenced by sample size than other fit statistics, such as RMSEA (Fan & Sivo, 2007). General guidelines suggest that a RMSEA less than .08 indicates acceptable fit (Hu & Bentler, 1999) and a CFI or TFI greater than .95 indicates excellent model fit (Hu & Bentler, 1999), although some consider a CFI or TFI greater than .90 (Browne & Cudeck, 1993) or a Gamma-hat greater than or equal to .90 (Marsh, Hau, & Wen, 2004) to indicate acceptable model fit. The chi-square (χ^2) values were used as well, but they were interpreted in respect to the other fit statistics as χ^2 values tend to be overly sensitive to minor differences when samples are large, e.g., $n > 200$ (Little, 2013).

Three difference tests were used to determine if the model fit was similar between configural and metric, and metric and scalar invariance models: (a) a chi-square difference test (χ^2), (b) CFI difference test (CFI), and (c) Gamma-hat difference test (Gamma-hat). A non-significant χ^2 test, $p > .05$, suggests that the more constrained model fits the data nearly as well as the freer model and provides support for invariance. The χ^2 may be sensitive to trivial differences if the sample size is larger and other statistics of model fit are needed to determine if invariance models are statistically similar (Brannick, 1995; Kelloway, 1995). A

CFI less than or equal to $-.010$ (Cheung & Rensvold, 2002) and a Gamma-hat less than or equal to $-.008$ (Chen, 2007) indicate similarity in fit between. Moreover, it is suggested that multiple differences in fit statistics are utilized and invariance is supported when the majority of the comparisons indicate similar model fit (Vandenberg & Lance, 2000).

Results

Scale and item means and scale reliability coefficients for each time point are provided by rater at each time point in Table 4. Within raters, mean scores remained consistent over time. However, scale scores were different by reporter, as parents gave more positive reports of their parenting than their children. Parents rated themselves higher on Positive Parenting, and youth rated their parents higher on Inconsistent Discipline and Poor Supervision. Scale reliability was consistent over time and across raters, but the α 's for the youth-reported Inconsistent Discipline and Poor Supervision were low ($< .6$).

The correlations among APQ-9 scales are presented in Table 5. The APQ-9 scales intercorrelations within raters were low to moderate. For parent-rated scales at each time point, correlations among the three APQ-9 scales were statistically significant, in the expected direction, and ranged from .15 to .30. For youth reported scales, correlations were in the expected direction, but the correlations between Positive Parenting and Inconsistent Discipline were non-significant at the last two time points. Across raters, concurrent correlations between parent and youth report on the same scales ranged from low to moderate in strength (range .09 to .32). All of these cross-rater correlations were statistically significant with the exception for Inconsistent Discipline at the end of ninth grade.

Parent Longitudinal Invariance

The results of the configural, metric, and scalar longitudinal invariance tests for parent and youth reports on the APQ-9 are located in Table 6. Whereas χ^2 's indicated significant global model misfit, the CFI, TLI, and Gamma-hat indicated adequate fit for all models to the data (Browne & Cudeck, 1993; Marsh, Hau, & Wen, 2004). The comparison between the configural and metric models yielded a non-significant χ^2 , CFI, and Gamma-hat. The comparison between the metric and scalar models yielded a significant χ^2 , but CFI and Gamma-hat indicated similar model fit. Longitudinal scalar invariance was supported for the parent-report APQ-9 scales. The standardized factor loadings in the scalar invariance model ranged from .753 to .908 for Positive Parenting, .531 to .760 for Inconsistent Discipline, and .512 to .761 for Poor Supervision.

Youth Longitudinal Invariance

All of the longitudinal invariance models for youth showed statistically significant χ^2 values of global model fit, but the CFI, TLI, and Gamma-hat values suggested adequate fit to the data (Browne & Cudeck, 1993; Marsh, Hau, & Wen, 2004). Both the comparison between the configural and metric and between the metric and scalar invariance models yielded non-significant χ^2 and CFI and Gamma-hat below cutoffs for evidence of non-invariance. Longitudinal scalar invariance was supported for the youth-report APQ-9 scales. In the scalar invariance model, standardized factor loadings ranged from .708 to .830 for Positive Parenting, .504 to .556 for Inconsistent discipline, and .391 to .720 for Poor Supervision. The lowest loadings were for the item "Your parent(s) do not know the friends you are with," reflecting the fact that youth answers to this item were weakly correlated with their answers to the two other items in the scale.

Multi-group Longitudinal Invariance

Table 7 contains the results of the invariance tests across parents and youth report. The comparison between the configural (Table 8 for coefficients) and metric (Table 9 for coefficients) models yielded a significant χ^2 , but CFI and Gamma-hat were below cut-off values for non-invariance. The comparison between the metric and scalar models yielded a significant χ^2 , and CFI, and Gamma-hat above the non-invariance cut-off. Multi-group longitudinal invariance was unsupported.

Based on evidence of non-invariance between the metric and scalar invariance models, constraints were released at the mean intercepts for the multi-group scalar invariance test. Constraints were released individually for each item across groups and the χ^2 was calculated. After each constraint was released, items were rank ordered from largest to smallest χ^2 . Constraints were then released from the scalar model in the rank order until a pattern of fixed and freely estimated parameters yielded an invariant model. The invariant model included intercept constraints released on 5 of 9 items across parents and youth for Compliments, Threat, Early Out, No Note, and Friends (Table 10 for coefficients). Due to the majority of the items being released to achieve invariance, the APQ-9 was determined to be scalar non-invariant across raters.

The APQ-9 Positive Parenting scale item intercept for Compliments was greater for the youth than the parents at the middle and end of eighth grade (.121 vs. -.079). As reflected in the pattern of means shown in Table 4, Youth reporters, more so than parents, may be more likely to report receiving compliments than to receive the two other forms of positive support. In other words, at a given level of response to the two other items in the scale, youth were more likely to give a higher response to the Compliment item than were parents. The Inconsistent Discipline scale item intercept for Threat was substantially lower for youth than parents (-.198 vs. -.033) whereas the item intercept for Early Out was greater for youth (.382 vs. .179). Youth may have difficulty discerning what a poor follow through with consequences would be, but they may be better able to grasp when they can manipulate their parent to discontinue punishments. The Poor Supervision scale item intercept for No Note was substantially lower for youth than parents (-.166 vs. .063) whereas the item intercept for Friends was greater for youth (.817 vs. .240). This, again, points to a difference between reporters in the relative likelihoods of giving affirmative answers to these items.

Within each group the means for latent factors remained consistent over time, which matches the pattern with the manifest scores. Parents' Positive Parenting latent scores were consistent at each time point (range = 4.30 to 4.37), as were their Poor Supervision scores (range = 1.61 to 1.67). Similarly, youth had stable Positive Parenting (range = 3.58 to 3.69) and Poor Supervision scores (range = 1.88 to 1.92). However, parents' scores for Inconsistent Discipline varied slightly (range = 2.14 to 2.49), whereas the youth's scores remained more stable (range = 2.42 to 2.52). Whereas the factor means from the partially invariant model show differences in means across reporters that seem to indicate more positive reports of parenting by parents than youth, the degree of scalar non-invariance suggests limitations on meaningful comparisons of factor means. Given that intercepts were freed for two indicators of both the Inconsistent Discipline and Poor Supervision factors, reporter differences in means for these factors are entirely a function of the means of the one item for which the intercept was constrained to equality.

Discussion

Youth and parent longitudinal metric and scalar invariance for the APQ-9 were supported when tested separately. This indicates that the youth and parent forms of the APQ-9 can be used independently to compare mean changes in these parenting practices across the transition from middle school to high school. There was some slight variability on the scores for each scale, but in general, means of manifest scales and on latent factors for both parents and youth showed little change across the youth transition from middle school to high school.

However, this study's results do call into question the practice of comparing APQ-9 scale means across parents and youth. In the multi-group models in which invariance across reporters was tested, the findings support configural and metric invariance, but fail to support scalar invariance. Adequate fit for a partially scalar invariant model required non-invariance in the majority of items representing inconsistent discipline and poor supervision. This finding implies that scale scores may have different relationships to underlying items for parents and their children and that comparison of scale means across reporters may not

be meaningful. For instance, the item means for Inconsistent Discipline indicate that youth gave more affirmative answers to the Early Out item relative to their answers to the Threat item compared to the relative likelihood of affirmative answers to these items for parents.

It may be that youth are more aware of the rewarding aspect of praise for task completion than parents. Youth may also be more perceptive of being allowed to escape negative consequences early, or parents may be hesitant to admit shortcomings with punishment consistency. The difference between youth and parents for the failure to leave a note may reflect the influence of technology, such that parents may consider leaving a note as leaving a handwritten note, whereas youth may consider sending electronic transmissions (e.g., text messages) as notes. Additionally, the parent knowledge of friends item may reflect that parents have comparatively less knowledge of their child's friends, whereas youth are more likely to expect parents to know less about their peer group. The divergence between youth and parent scores' scaling and item ratings may be expected due to changing parent-child interactions and perceptions of each other during the transition from middle school to high school (Granic & Patterson, 2006).

It is well-established that during early adolescence youth-parent cross-informant correspondence on behavioral questionnaires would be low (Tein, Roosa, & Michaels, 1994) and this study provides similarly low correlations across scales (e.g., Fleming et al., 2015). However, studies such as this one may shed light on to other aspects of correspondence and non-correspondence on measures during transitional periods. The metric invariance may indicate that parents and children may conceptualize the constructs of Positive Parenting, Inconsistent Discipline, and Poor Supervision as related to same respective activities, but due to differences in perception, each construct may be thought of as lying within different rater-contingent continua. Parents and youth conceptualize Positive Parenting, Inconsistent Discipline, and Poor Supervision as related to the same activities, but their differences in perception may lead to different scoring structures. That is, the APQ-9 measures the same constructs with parent and youth report, but the conceptual space between each anchor point on the scale differs between the groups. Still, to adequately assess how youth and parents are similar and different in their appraisal of aspects of parenting may require more items.

Additionally, the modest factor loadings ($< .70$) on the Inconsistent Discipline and Poor Supervision scales for both youth and parents could indicate the items had limited representation of the respective constructs. This is related to the low reliabilities for the Inconsistent Discipline and Poor Supervision scales. The accepted standards for research instrument reliability ($\alpha = .80$; Nunnally, 1978) is more stringent than for clinical instrument reliability (between .40 to .59 as fair, .60 to .74 as good, and .75 to 1.00 as excellent; Cicchetti, 1994). Parent and youth reliability both were below the research standard, but parent reliability was in the good clinical reliability range, whereas youth reliability was in the fair range. Without items that load more strongly on to the Inconsistent Discipline and Poor Supervision constructs, it may be difficult to assure measurement precision for either intervention investigations or practical assessment. However, it may be fruitful to understand which item or items this is due to because either replacing or expanding the number of items may improve construct representation. For example, the knowledge of friends item on the Poor Supervision scale may be particularly problematic.

The relation of the item to the construct varied widely across raters, as indicated by the standardized parameter estimates, and the mean intercept divergence pointed to potentially widely different interpretations of the item across reporters.

Parent- and child-rater perspective can account for some of the differences in item ratings (Janssens et al., 2015; Kerr & Stattin, 2000; Kerr et al., 2010), but it should be noted that not all of the items refer specifically to parent behavior. For example, the Poor Supervision scale items refer primarily to child behavior. These items appear to take an inferential approach to assessing parent behaviors, such that, the presence of child misbehavior is attributed to a deficit in parenting rather than any other number of potential factors. Future research should consider the degree to which altering the wording of the items may make responding to and interpretation of the items more consistent.

In general, it may be expected that ratings given by parents and children would become invariant because data collection spanned the transition from middle school to high school. It has been shown that the nature of the parent-child relationship changes during this critical developmental period (Granic & Patterson, 2006). Adolescents often experience increased autonomy as they transition into high school, which typically coincides with decreased direct parental involvement. These types of changes could impact how each rater endorses parenting items. Specifically, parenting measures given during adolescence may be more likely to reflect the types of information parents are able to gather about their children's behavior and what information children are willing to disclose to their parent or other adults (Kerr & Stattin, 2000; Kerr et al., 2010).

Limitations and Future Directions

There are limitations to this study which need to be addressed. The participants were targeted for a prevention program. There were no treatment effects across study conditions, but parents and youth may have responded differently than if they were not in an intervention study. Recent findings indicate that invariance between parenting constructs may be moderated by parent gender (i.e., mother or father; Janssens et al., 2015). Most of the parent participants in our study were mothers, and we were not able to examine differences between mother and father responses to the APQ. More study is warranted to determine how and how come parent gender may alter youth ratings of parenting. Additionally, the participants were from at-risk schools in an urban area with predominantly low socioeconomic status (SES) families. Data should be collected to determine if the study findings are consistent across parents and youth throughout a range of communities and family SES. The families also belonged to communities in the Northwest United States. There is variation within the United States as to when students transition out of middle school. For instance, middle schools that house sixth to eighth grade students may allow for students to begin developing independence earlier than middle schools that have seventh and eighth grade students.

Another limitation is that the age range was limited in this study. The APQ was developed for and normed on parents and children from 6 to 17 years of age (Frick et al., 1999; Shelton et al., 1996), and the initial psychometric studies for the APQ-9 were conducted with children 4 to 18 years of age (Elgar et al., 2007). It would be beneficial to replicate this

study across different age groups or developmental ranges. A limitation with the APQ-9 supported in this study was the low reliabilities for the Inconsistent Discipline and Poor Supervision scales, especially for the youth version. It is accepted that short or alternative forms of an assessment typically have lower reliability scores than the full forms, and that including more items may improve a scale's reliability (Henson, 2001). This study could have been strengthened by comparing the APQ-9 scales and items to the full APQ scales and items. Moreover, this study only used the APQ-9 and studies with other parenting scales may illuminate how parents and children perceive parenting differently.

From the results gathered in this study, it may be reasonable to surmise that differences in APQ-9 scale scores may be indicative of difference in parent and youth point of view. Both researchers and practitioners could view score differences as a reflection of the developmentally typical changes in the parent-child relationship, rather than an issue of accuracy. Moreover, the low reliabilities and tendency for youth responses to have lower correspondence to the scale constructs may indicate that the youth reports could be better understood using the full form of the APQ, rather than the APQ-9.

Acknowledgments

The project described was supported by National Institute on Drug Abuse Grant # 1R01DA025651 to Boys Town National Research Institute for Child and Family Studies, and analyses and manuscript preparation by the Institute of Education Sciences, U.S. Department of Education, through Grant R324B110001 to the University of Nebraska-Lincoln.

References

- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232. [PubMed: 3562706]
- Barry TD, Lochman JE, Fite PJ, Wells KC, Colder CR. The influence of neighborhood characteristics and parenting practices on academic problems and aggression outcomes among moderately to highly aggressive children. *Journal of Community Psychology*. 2012; 40:372–379. DOI: 10.1002/jcop.20514
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Boyle MH, Jadad AR, Allnut DR. Lessons from large trials: The MTA study as a model for evaluating the treatment of childhood psychiatric disorder. *Canadian Journal of Psychiatry*. 1999; 44:991–998. [PubMed: 10637678]
- Brannick MT. Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*. 1995; 16:201–213.
- Brown, TA. *Confirmatory factor analysis for applied research*. 2. New York: Guilford Press; 2015.
- Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*. 2007; 14:464–504.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–290.
- De Los Reyes A, Goodman KL, Kliewer W, Reid-Quinones K. The longitudinal consistency of mother and child reporting discrepancies of parental monitoring and their ability to predict child delinquent behaviors two years later. *Journal of Youth and Adolescence*. 2010; 39:1417–1430. [PubMed: 20020188]

- Dimitrov DM. Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*. 2010; 43:121–149.
- Elgar FJ, Waschbusch DA, Dadds MR, Sigvaldason N. Development and validation of a short form of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies*. 2007; 16:243–259.
- Elgar FJ, Waschbusch DA, McGrath PJ, Stewart SH, Curtis LJ. Temporal relations in daily-reported maternal mood and disruptive child behavior. *Journal of Abnormal Child Psychology*. 2004; 32:237–247. [PubMed: 15228173]
- Essau CA, Sasagawa S, Frick PJ. Psychometric properties of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies*. 2006; 15:595–614.
- Fan X, Sivo SA. Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*. 2007; 42:509–529.
- Fleming CB, Mason WA, Thompson RW, Haggerty KP, Gross TJ. Child and parent report of parenting as predictors of substance use and suspensions from school. *The Journal of Early Adolescence*. 2015 Advanced On-line.
- Frick PJ, Christian RE, Wootton JM. Age trends in the association between parenting practices and conduct problems. *Behavior Modification*. 1999; 23:106–128.
- Graham, John W. Missing data analysis: Making it work in the real world. *Annual Review of Psychology*. 2009; 60:549–576. DOI: 10.1146/annurev.psych.58.110405.085530
- Granic I, Patterson GR. Toward a comprehensive model of antisocial development: a dynamic systems approach. *Psychological Review*. 2006; 113:101–131. [PubMed: 16478303]
- Gresham FM, Cook CR, Collins T, Dart E, Rasetshwane K, Truelson E, Grant S. Developing a change-sensitive brief behavior rating scale as a progress monitoring tool for social behavior: An example using the Social Skills Rating System—Teacher Form. *School Psychology Review*. 2010; 39:364–379.
- Hawes DJ, Dadds MR. Assessing parenting practices through parent-report and direct observation during parent-training. *Journal of Child and Family Studies*. 2006; 15:554–567.
- Henson RK. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*. 2001; 34:177–189.
- Hinshaw SP, Owens EB, Wells KC, Kraemer HC, Abikoff HB, Arnold LE, Wigal T. Family processes and treatment outcome in the MTA: Negative/ineffective parenting practices in relation to multimodal treatment. *Journal of Abnormal Child Psychology*. 2000; 28:555–568. [PubMed: 11104317]
- Hoeve M, Dubas JS, Eichelsheim VI, van der Laan PH, Smeenk W, Gerris JR. The relationship between parenting and delinquency: A meta-analysis. *Journal of Abnormal Child Psychology*. 2009; 37:749–775. [PubMed: 19263213]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Jacob T, Windle M. Family assessment: Instrument dimensionality and correspondence across family reporters. *Journal of Family Psychology*. 1999; 13:339–354.
- Janssens A, Goossens L, Van Den Noortgate W, Colpin H, Verschueren K, Van Leeuwen K. Parents' and adolescents' perspectives on parenting: Evaluating conceptual structure, measurement invariance, and criterion validity. *Assessment*. 2015; 22:473–489. DOI: 10.1177/1073191114550477 [PubMed: 25225229]
- Kelloway K. Structural equation modelling in perspective. *Journal of Organizational Behavior* (1986–1998). 1995; 16:215–224.
- Kerr M, Stattin H. What parents know, how they know it, and several forms of adolescent adjustment: Further support for a reinterpretation of monitoring. *Developmental Psychology*. 2000; 36:366–380. [PubMed: 10830980]
- Kerr M, Stattin H, Burk WJ. A reinterpretation of parental monitoring in longitudinal perspective. *Journal of Research on Adolescence*. 2010; 20:39–64.
- Kline, RB. Principles and practice of structural equation modeling. 3. New York: Guilford Press; 2010.
- Little, TD. Longitudinal structural equation modeling. New York: Guilford Press; 2013.

- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341.
- Mason WA, Fleming CB, Ringle JL, Thompson RW, Haggerty KP, Snyder JJ. Reducing Risks for Problem Behaviors During the High School Transition: Proximal Outcomes in the Common Sense Parenting Trial. *Journal of Child and Family Studies*. 2014; 24:2568–2578. DOI: 10.1007/s10826-014-0059-5 [PubMed: 26508822]
- Mason WA, Fleming CB, Ringle JL, Thompson RW, Haggerty KP, Snyder JJ. Reducing risks for problem behaviors during the high school transition: Proximal outcomes in the Common Sense Parenting trial. *Journal of Child and Family Studies*. 2015; 24:2568–2578. [PubMed: 26508822]
- Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*. 2010; 3:111–130.
- Molinuevo B, Pardo Y, Torrubia R. Psychometric analysis of the Catalan version of the Alabama Parenting Questionnaire (APQ) in a community sample. *The Spanish Journal of Psychology*. 2011; 14:944–955. [PubMed: 22059338]
- Muthén, B., Muthén, L. Mplus user's guide. 7. Los Angeles, CA: Muthén & Muthén; 1998–2013.
- Nunnally, JC. Psychometric theory. 2. New York, NY: McGraw-Hill; 1978.
- Pasch KE, Stigler MH, Perry CL, Komro KA. Parents' and children's self-report of parenting factors: How much do they agree and which is more strongly associated with early adolescent alcohol use? *Health Education Journal*. 2010; 69:31–42. [PubMed: 20563227]
- Schafer JL, Graham JW. Missing data: Out view of the state of the art. *Psychological Methods*. 2002; 7:147–177. DOI: 10.1037/1082-989X.7.2.147 [PubMed: 12090408]
- Scott S, Briskman J, Dadds MR. Measuring parenting in community and public health research using brief child and parent reports. *Journal of Child and Family Studies*. 2011; 20:343–352.
- Shelton KK, Frick PJ, Wootton J. Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*. 1996; 25:317–329.
- Steiger, JH., Lind, JC. Statistically-based tests for the number of common factors. Paper presented at the annual spring meeting of the Psychometric Society; Iowa City, IA. 1980.
- Tein JY, Roosa MW, Michaels M. Agreement between parent and child reports on parental behaviors. *Journal of Marriage and the Family*. 1994; 56:341–355. DOI: 10.2307/353104
- Tucker L, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*. 1973; 38:1–10.
- Tyler KA, Johnson KA, Brownridge DA. A longitudinal study of the effects of child maltreatment on later outcomes among high-risk adolescents. *Journal of Youth and Adolescence*. 2008; 37:506–521.
- van de Schoot R, Lugtig P, Hox J. A checklist for testing measurement invariance. *European Journal of Developmental Psychology*. 2012; 9:486–492.
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–70.
- Volpe RJ, Gadow KD. Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review*. 2010; 39:350–363.

Table 1

Parent Demographic Information at Baseline and Follow-up

	<u>Baseline</u>	<u>Follow-up</u>
<i>n</i>	321	303
Age <i>M(SD)</i>	40.7 years (7.69)	
Parent Type		
Mother	68%	69%
Father	15%	13%
Other	17%	18%
Race		
Caucasian	52%	50%
African American	26%	28%
Asian American	4%	6%
Pacific Islander	4%	5%
Native American	1%	2%
Mixed or "Other"	13%	9%
Hispanic	12%	14%
Highest Level of Education		
High School Diploma or Equivalent	50%	28%
Some College	37%	58%
Bachelor's or more advanced degree	13%	14%
Household incomes below \$24,000	41%	40%
Receive Food Stamps	60%	46%
Employment		
Full time	43%	50%
Part time	14%	9%
Unemployed	15%	9%
Unknown	27%	32%

Table 2

Youth Demographic Information at Baseline and Follow-up

	Baseline	Follow-up
<i>n</i>	321	307
Age <i>M(SD)</i>	13.46 years (0.53)	14.93 years (0.55)
Gender		
Boy	47%	48%
Girl	53%	52%
Race		
Caucasian	39%	42%
African American	35%	33%
Asian American	8%	6%
Pacific Islander	4%	6%
Native American	4%	3%
Mixed or "Other"	10%	10%
Hispanic	11%	9%

Table 3

Alabama Parenting Questionnaire-9 Scales and Items by Rater

Scale/Item	Youth	Parent
Positive Parenting	"How often do:..."	"How often do:..."
Good Job	Your parent(s) tell you that you are doing a good job.	You let your child know when he/she is doing a good job with something.
Compliment	Your parent(s) compliment you when you have done something well.	You compliment your child after he/she has done something well.
Praise	Your parent(s) praise you for behaving well.	You praise your child if he/she behaves well.
Inconsistent Discipline	"How often do:..."	"How often do:..."
Threat	Your parent(s) threaten to punish you and then do not do it.	You threaten to punish your child and then do not actually punish him/her.
Early Out	Your parent(s) let you out of a punishment early (like lift restrictions earlier than they originally said.)	You let your child out of a punishment early (like lift restrictions earlier than you originally said).
Talk Out	You talk your parent(s) out of punishing you after you have done something wrong.	Your child talks you out of being punished after he/she has done something.
Poor Supervision	"How often do:..."	"How often do:..."
No Note	You fail to leave a note or let your parent(s) know where you are going.	Your child fails to leave a note or to let you know where he/she is going.
Stay Out	You stay out in the evening past the time you are supposed to be home.	Your child stays out in the evening after the time he/she is supposed to be home.
Friends	Your parent(s) do not know the friends you are with.	Your child is out with friends you don't know.

Table 4

APQ-9 Scale and Item Means and Standard Deviations with Scale Reliability, by Reporter.

Scale ^a	Middle of 8 th Grade			End of 8 th Grade			End of 9 th Grade		
Parent	<i>M</i> (<i>SD</i>)	<i>α</i>		<i>M</i> (<i>SD</i>)	<i>α</i>		<i>M</i> (<i>SD</i>)	<i>α</i>	
APQ-PP	13.1 (1.8)	.81		13.1 (2.0)	.89		12.9 (2.0)	.83	
Good Job	4.4 (0.7)			4.4 (0.7)			4.3 (0.7)		
Compliment	4.4 (0.7)			4.4 (0.7)			4.3 (0.8)		
Praise	4.3 (0.7)			4.3 (0.8)			4.3 (0.8)		
APQ-ID	7.4 (2.3)	.61		6.9 (2.3)	.64		6.4 (2.2)	.66	
Threat	2.7 (1.0)			2.5 (1.0)			2.2 (1.0)		
Early Out	2.6 (1.0)			2.4 (1.0)			2.3 (0.9)		
Talk Out	2.2 (1.1)			2.0 (1.0)			1.9 (1.1)		
APQ-PS	4.9 (2.0)	.64		4.9 (2.0)	.65		5.0 (2.1)	.70	
No Note	2.0 (1.0)			1.9 (1.0)			1.9 (1.1)		
Stay Out	1.4 (0.8)			1.4 (0.7)			1.5 (0.8)		
Friends	1.5 (0.8)			1.6 (0.8)			1.6 (0.7)		
Youth									
APQ-PP	11.2 (2.9)	.81		11.0 (2.7)	.81		10.7 (2.9)	.84	
Good Job	3.8 (1.0)			3.7 (1.0)			3.6 (1.0)		
Compliment	3.9 (1.1)			3.8 (1.1)			3.7 (1.1)		
Praise	3.5 (1.2)			3.4 (1.2)			3.4 (1.2)		
APQ-ID	7.3 (2.5)	.56		7.4 (2.3)	.53		7.4 (2.4)	.57	
Threat	2.4 (1.2)			2.5 (1.1)			2.4 (1.1)		
Early Out	2.8 (1.1)			2.7 (2.7)			2.8 (1.1)		
Talk Out	2.1 (1.2)			2.2 (2.2)			2.2 (1.1)		
APQ-PS	6.1 (2.5)	.50		6.1 (2.4)	.57		6.0 (2.4)	.57	
No Note	2.0 (1.2)			2.1 (1.2)			1.9 (1.1)		
Stay Out	1.6 (1.0)			1.7 (1.0)			1.8 (1.0)		
Friends	2.4 (1.3)			2.3 (1.2)			2.3 (1.2)		

Note. APQ-PP = Positive Parenting. APQ-ID = Inconsistent Discipline. APQ-PS = Poor Supervision.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Wording for the scale items is provided for the parent APQ-9.

Good Job = You let your child know when he/she is doing a good job with something. Compliment = You compliment your child after he/she has done something well. Praise = You praise your child if he/she behaves well. Threat = You threaten to punish your child and then do not actually punish him/her. Early Out = You let your child out of a punishment early (like lift restrictions earlier than you originally said). Talk Out = Your child talks you out of being punished after he/she has done something. No Note = Your child fails to leave a note or to let you know where he/she is going. Stay Out = Your child stays out in the evening after the time he/she is supposed to be home. Friends = Your child is out with friends you don't know.

Table 5**Bivariate Correlations of APQ-9 Scales for Parents and Youth**

	Middle of 8 th Grade			End of 8 th Grade			End of 9 th Grade		
	1	2	3	1	2	3	1	2	3
n = 286									
1. APQ-9-PP	.138 *	-.149*	-.206**	.188 **	-.212**	-.240**	.214 **	-.178**	-.247**
2. APQ-9-ID	-.211**	.228 **	.290**	-.045	.261 **	.363**	-.038	.092	.298**
3. APQ-9-PS	-.259**	.341**	.320 **	-.214**	.271**	.266 **	-.143*	.305**	.260 **

Notes: Bold diagonal coefficients are correlations between parent and youth scale scores. Parent scales scores intercorrelations are above the diagonals. Youth scales scores intercorrelations are below the diagonals. APQ-9-PP = Positive Parenting scale, APQ-9-ID = Inconsistent Discipline scale, APQ-9-PS = Poor Supervision scale.

* p < .05.

** p < .01.

*** p < .001.

Table 6
Parent and Youth Longitudinal Invariance Model Fit Test Statistics and Fit Indices and their Differences across Invariance Tests

Model	Model Fit					Difference in Model Fit ^d				
	χ^2	df	RMSEA [90% CI]	CFI	TLI	Gamma-hat	χ^2	df	CFI	Gamma-hat
Parent										
Configural	353.1***	261	.033 [.024, .042]	.972	.962	.979				
Metric	357.4***	273	.031 [.021, .040]	.974	.966	.981	4.3	12	.002	-.002
Scalar	401.0***	285	.036 [.027, .043]	.964	.956	.973	43.6***	12	-.010	-.008
Youth										
Configural	392.5***	261	.040 [.031, .047]	.951	.934	.970				
Metric	408.0***	273	.039 [.031, .047]	.950	.935	.970	15.5	12	-.001	<-.001
Scalar	428.1***	285	.040 [.032, .047]	.947	.934	.967	20.1	12	-.003	-.003

Notes.

^a Difference of Model Fit and Fit Indices are Metric vs. Configural, and Scalar vs. Metric.

* p < .05.

**

p < .01.

p < .001

Table 7

Longitudinal Invariance Model Fit Test Statistics and Fit Indices and their Differences across Invariance Tests Across Parents and Youth

Model	Model Fit					Difference in Model Fit ^a				
	χ^2	df	RMSEA [90% CI]	CFI	TLI	Gamma-hat	χ^2	df	CFI	Gamma-hat
Configural	745.6	522	.037 [.030, .042]	.962	.949	.974				
Metric	799.8	552	.037 [.032, .043]	.958	.947	.973	54.2**	30	-.004	-.001
Scalar	1051.0	582	.050 [.045, .055]	.921	.904	.949	251.2***	30	-.037	-.024
Scalar - Invariant ^b	868.5	577	.040 [.034, .045]	.951	.940	.967	68.7***	25	-.007	-.006

Notes.^aDifference of Model Fit and Fit Indices are Metric vs. Configural, and Scalar vs. Metric.^bPartial invariant model: Parameter constraints released for youth intercepts for Compliment, Threat, Early Out, No Note, Friends.

Compliment = You compliment your child after he/she has done something well. Threat = You threaten to punish your child and then do not actually punish him/her. Early Out = You let your child out of a punishment early (like lift restrictions earlier than you originally said). No Note = Your child fails to leave a note or to let you know where he/she is going. Friends = Your child is out with friends you don't know.

* $p < .05$.** $p < .01$.*** $p < .001$.

Table 8

Multi-group Longitudinal Configural Invariance Test Parameter Estimates, Standardized Parameter Estimates, and Means Intercept Values.

Scale	Middle 8 th Grade		End 8 th Grade		End 9 th Grade	
	Parent Estimate ^a	Intercept	Youth Estimate ^a	Intercept	Parent Estimate ^a	Intercept
APQ-PP						
Good Job	.964 (.784)	.940 (.797)	.908 (.803)	.891 (.769)	.923 (.784)	.960 (.831)
	.180	.292	.411	.457	.359	.211
Compliment	1.01 (.765)	1.06 (.801)	1.05 (.921)	1.18 (.887)	1.05 (.817)	1.03 (.816)
	-.002	-.047	-.146	-.479	-.173	.009
Praise	1.03 (.764)	.995 (.716)	1.05 (.834)	.933 (.655)	1.03 (.744)	(.760)
	-.178	-.245	-.265	.022	-.186	-.220
APQ-ID						
Threat	.981 (.625)	1.21 (.647)	.936 (.492)	.978 (.492)	.983 (.631)	.951 (.514)
	.301	-.546	.342	.113	.054	.110
Early Out	1.19 (.731)	.801 (.439)	1.15 (.392)	.741 (.392)	1.14 (.802)	.887 (.479)
	-.392	.836	-.278	.854	-.110	.581
Talk Out	.833 (.475)	.992 (.534)	.915 (.686)	1.30 (.686)	.878 (.500)	1.16 (.653)
	.091	-.291	-.064	-.967	.057	-.691
APQ-PS						
No Note	1.15 (.586)	1.20 (.640)	1.07 (.578)	1.20 (.633)	1.14 (.626)	1.12 (.604)
	.120	-.414	.185	-.351	.044	-.310
Stay Out	.897 (.624)	1.30 (.793)	1.01 (.750)	1.30 (.794)	.994 (.704)	1.23 (.775)
	-.074	-.983	-.243	-.852	-.169	-.704
Friends	.954 (.659)	.494 (.234)	.923 (.592)	.540 (.284)	.869 (.684)	.650 (.328)
	-.046	1.397	.058	1.204	.125	1.014

Notes.

^aStandardized estimates in parentheses.

^aWording for the scale items is provided for the parent APQ-9. Good Job = You let your child know when he/she is doing a good job with something. Compliment = You compliment your child after he/she has done something well. Praise = You praise your child if he/she behaves well. Threat = You threaten to punish your child and then do not actually punish him/her. Early Out = You let your child out of a punishment early (like lift restrictions earlier than you originally said). Talk Out = Your child talks you out of being punished after he/she has done something. No Note = Your child fails to leave a note or to let you know where he/she is going. Stay Out = Your child stays out in the evening after the time he/she is supposed to be home. Friends = Your child is out with friends you don't know.

Table 9

Multi-group Longitudinal Metric Invariance Test^a Parameter Estimates, Standardized Parameter Estimates, and Means Intercept Values.

	APQ-PP			APQ-ID			APQ-PS		
	<u>Good Job</u>	<u>Compliment</u>	<u>Praise</u>	<u>Threat</u>	<u>Early Out</u>	<u>Talk Out</u>	<u>No Note</u>	<u>Stay Out</u>	<u>Friends</u>
Estimate ^b	.930	1.05	1.02	1.00	1.02	.977	1.16	1.05	.794
Mean Intercept ^d									
Mid 8 th Grade	.328/.329	-.202/-.007	-.126/-.322	.246/-.049	.021/.304	-.267/-.255	.110/-.318	-.324/-.473	.214/.791
End 8 th Grade	.313/.314	-.174/-.028	-.139/-.285	.187/.051	.022/.167	-.209/-.218	.048/-.226	-.317/-.461	.268/.687
End 9 th Grade	.330/.319	-.204/-.061	-.125/-.258	.011/-.017	.145/.255	-.155/-.238	.014/-.372	-.264/-.353	.251/.726

Notes.

^aFactor loadings constrained to equality across raters; standardized loadings available by request, from the first author.

^bParameter estimates are constrained to equality across parents and youth over each time point.

^cParent Standardized Estimates are to the left of the slash.

^dParent means intercepts are to the left of the slash.

^aWording for the scale items is provided for the parent APQ-9. Good Job = You let your child know when he/she is doing a good job with something. Compliment = You compliment your child after he/she has done something well. Praise = You praise your child if he/she behaves well. Threat = You threaten to punish your child and then do not actually punish him/her. Early Out = You let your child out of a punishment early (like lift restrictions earlier than you originally said). Talk Out = Your child talks you out of being punished after he/she has done something. No Note = Your child fails to leave a note or to let you know where he/she is going. Stay Out = Your child stays out in the evening after the time he/she is supposed to be home. Friends = Your child is out with friends you don't know.

Table 10

Multi-group Longitudinal Scalar Invariance Test^a Parameter Estimates, Standardized Parameter Estimates, and Means Intercept Values.

	APQ-PP			APQ-ID			APQ-PS		
	<u>Good Job</u>	<u>Compliment</u>	<u>Praise</u>	<u>Threat</u>	<u>Early Out</u>	<u>Talk Out</u>	<u>No Note</u>	<u>Stay Out</u>	<u>Friends</u>
Estimate ^b	.908	1.06	1.04	1.08	.972	.950	1.15	1.05	.796
Non-Invariant ^c									
Mean Intercept	.282	.044	-.326	-.009	.186	-.178	.116	-.310	.194
Invariant ^d									
Parent Mean Intercept	.433	-.195	-.237	-.038	.189	-.151	.062	-.302	.240
Youth Mean Intercept		.032		-.172	.380		-.167		.817

Notes.

^aScalar non-invariant and invariant model coefficients presented; multiple constraints released across parents and youth to reach invariance; standardized loadings available by request, from the first author.

^bParameter estimates are constrained to equality across parents and youth.

^cNon-invariant scalar model; means intercepts are constrained across parents and youth.

^dInvariant scalar model; means intercepts that are the same between parents and youth are not shown.

^aWording for the scale items is provided for the parent APQ-9. Good Job = You let your child know when he/she is doing a good job with something. Compliment = You compliment your child after he/she has done something well. Praise = You praise your child if he/she behaves well. Threat = You threaten to punish your child and then do not actually punish him/her. Early Out = You let your child out of a punishment early (like lift restrictions earlier than you originally said). Talk Out = Your child talks you out of being punished after he/she has done something. No Note = Your child fails to leave a note or to let you know where he/she is going. Stay Out = Your child stays out in the evening after the time he/she is supposed to be home. Friends = Your child is out with friends you don't know.