Author Manuscript

Assessment of Adolescents' Victimization, Aggression, and Problem Behaviors: Evaluation of the Problem Behavior Frequency Scale

Albert D. Farrell, **Terri N. Sullivan**, **Elizabeth A. Goncy**, **Anh-Thuy H. Le**
Virginia Commonwealth University

Abstract

This study evaluated the Problem Behavior Frequency Scale (PBFS), a self-report measure designed to assess adolescents' frequency of victimization, aggression, and other problem behaviors. Analyses were conducted on a sample of 5,532 adolescents from 37 schools at four sites. About half (49%) of participants were male; 48% self-identified as Black non-Hispanic; 21% as Hispanic, 18% as White Non-Hispanic. Adolescents completed the PBFS and measures of beliefs and values related to aggression, and delinquent peer associations at the start of the sixth grade and over two years later. Ratings of participants' behavior were also obtained from teachers on the Behavioral Assessment System for Children. Confirmatory factor analyses supported a seven-factor model that differentiated among three forms of aggression (physical, verbal, and relational), two forms of victimization (overt and relational), drug use, and other delinquent behavior. Support was found for strong measurement invariance across gender, sites, and time. The PBFS factors generally showed the expected pattern of correlations with teacher ratings of adolescents' behavior and self-report measures of relevant constructs.

Keywords

Assessment of Aggression; Assessment of Victimization; Assessment of Problem Behaviors in Adolescence; Measurement Invariance

In recent years increasing attention has focused on the study of aggression and victimization during adolescence. Researchers have conducted studies to estimate the prevalence of these constructs, determine their trajectories over time, identify related risk and protective factors, and evaluate the impact of a variety of approaches to prevention (United States Department of Health and Human Services, 2001). These efforts all have one thing in common – the need for carefully developed measures to provide a solid foundation for this work. This, in turn, requires the resolution of several important issues. These include determining the underlying structure of aggression and victimization, evaluating the value of different approaches to their measurement, and using appropriate methods to establish their psychometric properties. The current study evaluated the Problem Behavior Frequency Scale (PBFS). The PBFS was developed to provide a self-report measure of specific forms of

aggression (i.e., physical, verbal, and relational) and victimization (overt and relational), and related problem behaviors (i.e., drug use, and other delinquent behavior). The aims of this study were to evaluate the factor structure of the PBFS; determine its measurement invariance across gender, schools from different locations in the United States, and time points spanning the beginning and end of middle school; and evaluate its validity based on its relation to relevant teacher-and self-reported constructs.

Although researchers have used a variety of approaches, self-report is the most commonly used method to assess adolescents' aggression and victimization (Furlong, Sharkey, Felix, Tanigawa, & Green, 2010). Self-report has many advantages over other methods. Nearly all other methods, including teacher or parent ratings of adolescents' behavior, behavioral observations, and school archival records, assess adolescents' behavior in contexts where the presence of the observer (e.g., teacher, parent) makes the behaviors of interest less likely to occur (Barker, Tremblay, Nagin, Vitaro, & Lacourse, 2006). Ratings by teachers and parents may also be influenced by overall impressions of an adolescent and associated attributions (De Los Reyes & Kazdin, 2005). Archival data can provide useful information about school-level incidents but are limited to specific behaviors observed by school personnel, and there is variability in their definition and enforcement across schools and even among teachers within the same school. Peer nominations offer a useful perspective, but are dependent on the peers who provide nominations, which can make replication difficult (Solberg & Olweus, 2003). There may also be concerns that the nomination process may result in stigmatization, and youth may be reluctant to identify aggressive peers if they have concerns about confidentiality (Orpinas & Horne, 2006).

Self-report measures have clear strengths and weaknesses. They may be subject to social desirability, leading to underreporting of undesirable behaviors (DeVellis, 2011). Adolescents may also be limited in their ability to recall behaviors and experiences. Although some researchers have questioned the validity of self-report measures of problem behaviors (Farrington, 1999), others have argued that adolescents generally answer such questions truthfully (Thornberry & Krohn, 2000). Indeed, adolescents tend to report frequencies of problem behaviors that are higher than those based on ratings by parents or teachers (Rescorla et al., 2013). Self-report may also be a particularly valuable method of assessing victimization because others may not be aware of an adolescent's experiences (Desjardins, Thompson, Sukhawathanakul, Leadbeater, & MacDonald, 2013). Although self-report clearly has limitations, it has multiple advantages and is likely to continue to play an important role in research on aggression.

Measures of aggression have differed in how they represent the structure of aggression. A growing body of research has emphasized the importance of differentiating between direct and indirect forms of aggression (e.g., Card, Stucky, Sawalani, & Little, 2008). Direct aggression includes physical and verbal acts such as hitting, pushing, threatening physical force, and insults. Indirect aggression represents acts that do not directly confront the victim such as spreading rumors, damaging property or social exclusion. Card et al. (2008) noted that this distinction is supported by factor analyses of scales representing direct and indirect forms of aggression. Results of their meta-analysis found high correlations between measures of direct and indirect aggression (i.e., average $r = .76$), but differences in their

patterns of association with measures of adjustment. Whereas direct aggression was more strongly related to externalizing problems, poor peer relations, and low prosocial behavior, indirect aggression was more strongly related to internalizing problems and high prosocial behavior. They also found that these relations were moderated by several factors including age and gender.

As Card et al. (2008) themselves admitted, classifying measures as direct or indirect does not do full justice to the variety of frameworks researchers have used to develop measures of aggression. Some researchers have differentiated acts of aggression based on the intent of the aggressor. Physical aggression has been defined as the use or threat of physical force to cause harm or injure another person (Ostrov & Kamper, 2015), and relational or social aggression as acts that target the victim's relationships or social status (Galen & Underwood, 1997). Within this framework it is not clear where verbal acts of aggression such as insults might fit. Some measures have overt aggression scales that combine physical and verbal aggression (e.g., Rosen, Beron, & Underwood, 2013). Others have separate scales for verbal and physical aggression (e.g., Marsh et al., 2011). Whereas Card et al. (2008) categorized relational and social acts of aggression as indirect, others have challenged this noting that they may sometimes be direct (Ostrov & Kamper, 2015). Some researchers have created scales reflecting both the form and the motivation of the aggressor (i.e., reactive versus instrumental; Little, Henrich, Jones, & Hawley, 2003). Most recently researchers have identified *cyber aggression* or *cyber bullying* as an additional form of aggression, though others have argued that many such acts can be incorporated into existing frameworks (Mehari, Farrell, & Le, 2014). A further confusion in the measurement of aggression is use of the term bullying. Bullying has been defined to include acts of aggression that are repeated over time where the perpetrator has or is perceived to have power to enable them to exert control over the victim or limit the victim's ability to respond (Gladden, Vivolo-Kantor, Hamburger, & Lumpkin, 2014). Despite this distinction, items on many measures purported to assess bullying are very similar to those on other measures of aggression and do not typically incorporate elements of this definition (Furlong et al., 2010).

Although researchers have used similar frameworks to guide the development of victimization measures, there have been some key differences across studies, particularly in the treatment of verbal victimization. Rosen et al. (2013) conducted a confirmatory factor analysis on a version of the Revised Social Experience Questionnaire adapted for use with adolescents. Their results supported a two-factor model with separate factors representing overt and social victimization over a three-factor model that split overt aggression into separate factors for physical and verbal victimization. In contrast, Hunt, Peters, and Rapee (2012) found support for representing verbal and relational victimization items by a single factor in their analysis of a measure of bullying victimization. Finally, support has also been found for treating verbal victimization as a distinct factor (Marsh et al., 2011).

Researchers evaluating measures of adolescent aggression and victimization have become increasingly sophisticated in their application of statistical models relevant to evaluating measures of aggression and victimization. Response formats for many of these measures (e.g., *never, almost never, sometimes, almost all the time, all the time*) do not meet the equal-intervals assumption of conventional methods of factor analysis (Piquero, Macintosh, &

Hickman, 2000). This has led to increasing use of robust least squares estimators that are well suited for ordered categorical variables, and that can account for differences in the distances between ordinal categories and variations in severity across items (e.g., Rosen et al., 2013). There has also been increasing recognition of the importance of measurement invariance, or the degree to which measurement properties are consistent across groups and over time. Measurement invariance is critical for making meaningful comparisons over time or across groups (Widaman & Reise, 1997). Although such comparisons are often the focus of research on aggression and victimization, there have been few attempts to evaluate the measurement invariance of measures of these constructs (e.g., Marsee et al., 2011, Marsh et al., 2011, Rosen et al., 2013).

The PBFS was developed to provide a self-report measure to assess adolescents' frequency of victimization, aggression, and other domains of problem behaviors (e.g., drug use, nonviolent delinquency). It was originally designed to serve as an outcome measure for studies evaluating youth violence prevention programs (e.g.., Farrell, Kung, White, & Valois, 2000; Farrell, Meyer, Sullivan & Kung, 2003) and has been used in studies examining interrelations of problem behaviors in both cross sectional and longitudinal studies (e.g., Farrell, Sullivan, Esposito, Meyer & Valois, 2005), and relations between problem behaviors and related constructs (e.g., Farrell & Bruce, 1997; Farrell, Henry, Mays, & Schoeny, 2011). Since its initial development the PBFS has gone through several revisions to broaden its item pool to address a wider range of aggressive behaviors and victimization experiences (Sullivan, Farrell, & Kliewer, 2006). The PBFS currently includes items representing three forms of aggression (physical, verbal, and relational), two forms of victimization (overt and relational), drug use, and other delinquent behaviors.

The PBFS has several advantages over other self-report measures of adolescents' aggression and victimization. In contrast to measures that focus on either aggression or victimization, it addresses both. This is particularly important given the strong patterns of concurrent and longitudinal relations between perpetration and victimization (Bettencourt, Farrell, Liu, & Sullivan, 2013). The PBFS includes a minimum of six items for each form of aggression and victimization, which provides a clearer basis for examining the structure of aggression and victimization than measures that sample a limited aspect of these domains. In addition to aggression it includes items representing other forms of externalizing behavior including drug use and non-aggressive delinquent behavior, which may be of benefit to studies examining multiple outcomes. In contrast to measures that include items that resemble trait-like statements (e.g., "I am the kind of person who often fights with others", Little et al., 2003) or conditional statements (e.g., "When someone hurts me, I end up getting into a fight"; Marsee et al. 2011), PBFS items focus on the frequency of specific behaviors (e.g., hit or slapped someone, spread a false rumor about someone) that are often the target of interventions. The rating scale asks respondents to endorse the frequency of each item using an operationally-defined six-point frequency scale (e.g., *Never, 1–2 times, 3–5 times*), rather than more subjectively defined anchors (e.g., *never, almost never, sometimes, almost all the time, all the time*; Rosen et al., 2013). The PBFS also specifies the time frame (i.e., past 30 days), which can be important when interpreting scores or using it as a measure of change.

Despite its frequent use, there are no published studies of the psychometric properties of the PBFS other than statements about the internal consistency of individual scales and a factor analysis of an earlier version (Farrell et al., 2000). The current study took advantage of a large multisite data set to conduct a comprehensive analysis of the PBFS. A key purpose was to test competing models of its structure based on frameworks found in previous studies of the structure of aggression and victimization. We hypothesized that the items would best be represented by a seven-factor model with factors representing specific forms of aggression (physical, verbal, relational) and victimization (overt and relational), drug use, and delinquent behavior. This model was compared to models in which verbal aggression was combined with either relational or physical aggression; a model with a single factor representing all three forms of aggression; a model with a single problem behavior factor that incorporated aggression, drug use, and other delinquent behaviors; and a model that combined overt and relational victimization into a single victimization factor. Once the overall structure of the PBFS was determined, we conducted tests of measurement invariance to determine the consistency of the PBFS across gender, sites representing four cities in different locations across the U.S., and time (start of the sixth grade and over 2 years later). These included tests of both configural invariance (i.e., consistency of the overall structure of the scale across groups) and scalar (i.e., strong) invariance (i.e., the extent to which the scaling of the measure was consistent across groups).

We also evaluated the validity of the PBFS by examining its concurrent relation with teacher ratings of adolescents' problem behaviors on the Behavioral Assessment Scale for Children (BASC, Reynolds & Kamphaus, 1992) and scores on self-report measures of constructs related to adolescent problem behaviors. We hypothesized that compared to verbal and relational aggression, physical aggression would be more strongly related to student reports of their beliefs and values related to fighting, and to teacher ratings of their aggression. We further hypothesized that physical aggression, delinquent behavior, and drug use represented more extreme forms of problem behavior than verbal and relational aggression (Card et al., 2008) and would thus be more positively correlated with student reports of delinquent peer associations and teacher ratings of students' conduct problems, and more negatively correlated with teacher ratings of adaptive behavior. In contrast, we hypothesized that victimization factors would have weaker relations with adolescents' reports of their values and beliefs related to fighting and delinquent peer associations, and teacher ratings of students' aggression than would factors representing problem behaviors, but would be more strongly related to teacher ratings of students' anxiety and depression (Card et al., 2008). Finally, we hypothesized that overt victimization would be more strongly related to constructs associated with aggression because of its tendency to be related to perpetration of physical aggression (Bettencourt et al., 2013).

Method

Procedure and Participants

Secondary analyses were conducted on data from two cohorts of students recruited from 37 schools from four different sites as part of the Multisite Violence Prevention Project (MVPP; Henry, Farrell, & MVPP, 2004). These included 12 Chicago schools that served grades K-8,

eight middle schools in Durham, North Carolina, eight middle schools in Richmond, Virginia, and three urban and six rural middle schools in northeastern Georgia. All had high percentages of students from low-income families based on eligibility for the federal free or reduced price lunch program (42% to 96% across sites). MVPP was designed to evaluate the effects of a school-based universal violence prevention program and a selective family intervention. Two to three schools in each site were randomized to four conditions: universal intervention, selective intervention, combined (universal and selective) intervention, and no-intervention control. Details regarding its design, school recruitment, and community characteristics are reported by Henry et al., 2004. Details on measures are reported by Miller-Johnson, Sullivan, Simon, and MVPP (2004).

Participants were recruited in September of 2002 and 2003 from a random sample of approximately 100 students from the sixth grade rosters of each school or from all sixth graders in three Chicago schools that had less than 100 sixth graders. All procedures were approved by the institutional review boards at the participating universities and the Centers for Disease Control and Prevention. Parental permission and student assent were obtained from 5,625 of the 7,364 eligible students (76%). Research staff administered measures to students at each school using a computer-assisted interview. Data were collected from each cohort at the beginning and end of the sixth grade and at the end of the following two school years. The current study examined data from the first and last wave, which captured the beginning and end of middle school. Analyses were based on 5,532 students who participated in at least one of these waves. The sample was about evenly divided by sex (49% boys); 48% self-identified as Black Non-Hispanic, 21% as Hispanic, 18% as White Non-Hispanic, 8% endorsed more than one race. About half (48%) resided with both biological parents; 26% resided with a single parent.

The Problem Behavior Frequency Scale (PBFS)

The version of the PBFS used in MVPP was based on the measure developed by Farrell and colleagues (2000) and included scales assessing physical aggression (seven items), verbal aggression (six items), relational aggression (six items), drug use (six items), other forms of delinquent behavior (eight items), overt victimization (six items), and relational victimization (six items) (see Appendix). Many items on the Physical Aggression and Overt Victimization scales were based on the Youth Risk Behavior Survey (Kolbe, Kann, & Collins, 1993). Items on the Relational Aggression scale were similar to those on Crick and Grotpeter's (1995) measure of relational aggression, and the Relational Victimization items were based on the Social Experience Questionnaire (Crick & Grotpeter, 1996). The majority of items on the Nonphysical Aggression scale represented verbal aggression and were based on school observations and focus group discussions of interpersonal problem situations (Farrell, Ampy, & Meyer, 1998). Items on the Drug Use scale focused on gateway drugs (Kandel, 1975). Items on the Delinquent Behavior scale were based on items in Jessor and Iessor's (1977) Attitudes Toward Deviance Scale, supplemented with items representing nonviolent delinquent behaviors. Items on the Aggression, Drug Use, and Delinquent Behavior scales were preceded by the stem: "In the last 30 days, how many times have you?" Victimization Items were preceded by the stem: "In the last 30 days, how many times

has this happened to you?" All items were rated on a six-point frequency scale, 1 = *Never,* 2 = *1–2 times*, 3 = *3–5 times*, 4 = *6–9 times*, 5 = *10–19 times,* and 6 = *20 or more times.*

Measures of Participants' Beliefs, Values, and Peer Associations

The Individual Norms for Aggression and Alternatives scale is based on a measure by Henry, Cartland, Ruchross, and Monahan (2004). We used the Individual Norms for Aggression scale on which participants rated their approval of ten items representing aggressive responses to specific situations (e.g., "How would you feel if a kid hit someone who said something mean?"). Responses were rated on a three-point scale (i.e., *disapprove, neutral*, and *approve*). Alpha at Wave 2 was .84.

The Beliefs about Aggression and Alternatives scale (Farrell, Meyer, & White, 2001) asks participants to rate their agreement with items involving the use of aggression (e.g., "It's O.K. for me to hit someone to get them to do what I want.") on a four-point scale: 1 = *Strongly agree*, 2 = *Agree somewhat*, 3 = *Disagree somewhat*, 4 = *Strongly disagree*. We used the Beliefs Supporting Aggression scale which is based on the mean of seven items reversed-coded such that a high score reflects more favorable beliefs about aggression. The alpha at Wave 2 was .76.

The Delinquent Peer Associations scale asks adolescents how many of their close friends have engaged in ten delinquent behaviors (e.g., stolen property, used alcohol) in the last three months (Miller-Johnson et al., 2004). Items are rated on a five-point scale, ranging from 0 (none of them) to 4 (all of them) and are averaged to create an overall score reflecting involvement in delinquent activities by the respondent's close friends. The alpha at Wave 2 was .88.

The Goals and Strategies scale is based on a measure by Hopmeyer and Asher (1997). It describes four scenarios involving a potential conflict with a same-gender peer and asks respondents to rate their likelihood of using specific strategies to deal with them and their goals in each situation. We used scales representing participants' endorsement of revenge ("my goal would be trying to get back at him/her for what he/she just did") and maintaining relationship goals ("my goal would be trying to get along with this student"). Items are rated on a five-point scale ranging from 1 -*Really disagree* to 5 – *Really agree*. Scores are based on the average across scenarios with a high score reflecting a stronger endorsement of that goal. Alpha coefficients were .88 for both scales.

Teachers' Ratings of Students' Adjustment

Teachers rated students' behavior using the adolescent form of the Behavioral Assessment System for Children Teacher Rating Scale (BASC-TRS-A), a nationally normed measure of student behavior problems and assets (Reynolds & Kamphaus, 1992). The BASC-TRS-A was normed on a nationally-representative sample of 809 12 to 18 year old students from four regions of the U.S. The median internal consistency based on the normative sample was .90 with values for individual scales ranging from .77 to .95. Test-retest reliability over a 1-month period ranged from .75 to .89. Teachers rate each item on a four-point scale anchored by *Never* and *Almost Always*. The current study examined scores on the

Aggression, Conduct Disorder, Anxiety, and Depression scales and the Adaptive Behavior composite scale.

Analysis

We first conducted a content analysis to confirm the placement of items into scales. We then used Mplus 7.11 to test competing models of the factor structure of the PBFS; to evaluate measurement invariance across gender, sites, and time; and to examine relations between the PBFS factors and related constructs. Items were treated as ordered categorical variables through use of weighted least squares mean-and variance-adjusted estimators (WLSMV). This analysis is comparable to a graded response item-response theory model. Measurement parameters include factor loadings, and item thresholds, which represent the value of the underlying latent variable (e.g., physical aggression) at which there is a .50 probability of crossing into the next category on the rating scale (e.g., moving from *Never* to a higher category) (Embretson & Reise, 2000). Although participants rated each item on a six-point scale, initial analyses indicated that very few participants used the two highest rating points on the scale (i.e., on average 1.1% and 1.5% endorsed *10–19 times*, and 2.6% and 3.3% *endorsed 20 or more times* at Waves 1 and 2, respectively). These extremely low frequencies necessitated combining the three highest-order categories because the WLSMV estimator requires non-zero values in two-way frequency tables for each pair of variables.

Confirmatory factor analyses were used to compare the hypothesized seven-factor model of the PBFS to the five competing models. All models allowed the measurement error of each Wave 1 item to covary with the measurement error of that same item at Wave 2. This follows the recommendation of Pitts, West, and Tein (1996), who argued that there is strong theoretical justification for allowing errors of measurement for the same indicator to covary over time, noting that some portion of the measurement error associated with an individual indicator may represent systematic variance not shared with other indicators of the same underlying factor. The relative fit of each model was evaluated by comparing the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis fit index (TLI). The fit of each competing model was also directly compared to the seven-factor model using the difference test calculated by Mplus (see Asparouhov & Muthén, 2006) such that significant values indicated that the seven-factor fit model was a significant improvement over the competing model.

Once the structure of the PBFS was established, multiple group analyses were used to test measurement invariance across gender, site, and time. This involved comparing an unconstrained model that specified the same factor structure for each group (i.e., configural invariance) to a model that constrained factor loadings and thresholds for each factor to the same values across groups (i.e., scalar or strong factorial invariance), and a model that constrained factor loadings and thresholds to the same values across both groups and waves. We then tested additional constraints on the variances and covariances among factors within each wave. We followed the recommendations of Cheung and Rensvold (2002) who argued that the change in the CFI (i.e., ACFI) is a more appropriate test of measurement invariance than the chi-squared difference test because it is less sensitive to sample size. This was based on a Monte Carlo simulation that examined the performance of a variety of fit indices for

testing measurement invariance. In particular, they recommended that the null hypothesis of measurement invariance not be rejected if imposing higher degrees of measurement invariance does not reduce the CFI by .01 or more.

A final set of analyses examined the validity of the PBFS by testing hypotheses regarding patterns of correlations between PBFS factors and measures of related constructs based on self-reports and teacher ratings on the BASC at Wave 2. We examined relations at Wave 2 because teacher ratings at Wave 2 were collected near the end rather than beginning of a school year and were thus based on a larger sample of students' behavior. We also expected more variability in measures of problem behavior at Wave 2 when participants were older.

Results

Preliminary Analyses

Ten faculty and doctoral students on our research team independently reviewed and classified the PBFS items based on the seven hypothesized factors. Agreement averaged 89% across items. Results confirmed the original placement of items into scales with the exception of one item originally in the relational aggression scale (i.e., "Made fun of someone to make others laugh"), which was classified as verbal aggression by the research team. Another item from the physical victimization scale that was considered ambiguous (i.e., "a student asked you to fight) was excluded, as were three items on the delinquent behavior scale that represented school-specific status offenses (e.g., "cheated on a test"). A review of item-information curves, which indicate how well an item differentiates among individuals at different levels of the underlying latent variable, obtained from an initial analysis of the PBFS factors suggested eliminating four items that contributed limited information to the overall scores. These were an item on the Physical Aggression factor ("threatened to hurt a teacher"), two items on the Verbal Aggression factor ("gave mean looks to another student," "insulted someone's family"), and one item on the Relational Victimization factor ("had a kid tell lies about you to make other kids not like you anymore").

Structural Model of the PBFS

All of the models except the six-factor model that specified a single overall victimization factor (Model M5 in Table 1) and the two-factor model (Model M6) met the criteria of RMSEA values less than .04 and CFI and TLI values greater than .95 (see Table 1). The seven-factor model (Model M1) fit the data very well (i.e., RMSEA = .021, CFI = .971) and was a significant improvement over all five competing models based on the difference test (see Table 1). Although the seven-factor model specified separate factors for physical, verbal, and relational aggression, the Verbal Aggression factor was highly correlated with both the Physical Aggression (i.e., rs = .91 and .87 at waves 1 and 2, respectively) and Relational Aggression factors (i.e., rs = .85 and .79 at waves 1 and 2, respectively). The seven-factor model represented a significant improvement over six-factor models that combined the verbal aggression items with either the physical (Model M2) or relational aggression items (Model M3), however the improvement in fit was fairly small. The fit indices for these two six-factor models were fairly similar in value making it difficult to

favor one model over the other. Combining items representing all three forms of aggression into a single aggression factor (Model M4) resulted in a clear decrease in model fit. In conclusion, there was no clear basis for combining verbal aggression items with physical aggression items versus relational aggression items, and forming a single aggression factor from all of these items resulted in a clear decrease in model fit. Based on these findings subsequent analyses focused on the seven-factor model to determine if there was further support for differentiating among the three forms of aggression.

Two additional versions of the seven-factor model were analyzed to test several key assumptions. As previously noted, the initial seven-factor model allowed measurement errors of each item to correlate across waves. This was supported by analyses indicating that an alternative model that constrained these correlations to zero resulted in a significant decrease in model fit (see Model M1.1 in Table 1). Moreover, sensitivity analyses based on running all models included in this study without including correlated measurement errors did not result in any differences in the overall pattern of findings or conclusions. In several cases, however, excluding these parameters resulted in estimation problems. We also ran analyses to determine the extent to which item thresholds varied across items. One of the advantages of treating items as ordered categorical rather than simply averaging ratings across items is that it does not assume that values on the rating scale represent the same level of the underlying construct across items. For example, endorsing '1–2 times in the past 30 days' for the item *threatening someone with a weapon (gun, knife, club, etc.)* would be expected to represent a more serious indication of physical aggression than endorsing the same point on the rating scale for *shoved or pushed another kid*. We tested this assumption by comparing the fit of the original model (M1), which allowed thresholds to vary across items, to a model in which thresholds were constrained across items (i.e., values on the three threshold parameters did not differ across items). The constrained model (see Model M1.2) fit the data poorly, and resulted in a significant decrease in model fit compared to the original model. This supports the benefit of treating items as ordered categorical versus conventional approaches to measurement that make more restrictive assumptions. Based on these findings, all subsequent versions of the seven-factor model included serial correlations among measurement errors and allowed thresholds to vary across items.

Measurement Invariance Across Gender

Further analyses were conducted to examine measurement invariance across gender. Although the seven-factor model emerged as the best fitting model in the analysis of the total sample, it was possible that the factor structure might differ for boys and girls. This was examined by separate analyses by gender that compared the fit of the six models described in the preceding section (i.e., M1 to M6). The seven-factor model fit the data very well for both boys and girls (RMSEA = .022 and .019, CFI = .968 and .977, and TLI = .965 and .975, respectively) and significantly improved the fit relative to all other models based on the difference test (all ps < .001). This provided support for configural invariance across gender. Further analyses were conducted on the seven-factor model to test for scalar invariance. An initial multiple group model that specified the same seven-factor structure for boys and girls, but allowed parameter estimates to vary by gender fit the data very well (see Model G1 in Table 2). Model fit decreased very slightly (i.e., CFI = −.001) when factor loadings and

item thresholds were constrained across gender (Model G2), or across gender and waves (CFI = −.002; Model G3). This provided support for scalar or strong factorial invariance. In other words, the PBFS not only has the same factor structure for male and female adolescents, but it can be scored using the same loadings and item thresholds for male and female adolescents across both waves of data.

Invariance in the measurement structure of the PBFS provided a basis for examining differences in the means and patterns of relations among the seven factors for male and female adolescents. Gender differences in correlations among the seven factors were tested by constraining factor variances to 1 and covariances to the same values for boys and girls. This more restrictive model (Model G4) slightly improved the fit relative to the less restrictive model (Model G3). Gender differences in factor means at each wave were tested using the constraint function in Mplus to calculate an omnibus Wald test and follow-up tests using a per-test p-value of .004 to control for Type I error. Based on this criterion, boys reported higher levels of physical and verbal aggression, physical victimization, and delinquent behavior than did girls at both waves (see Figure 1). In contrast, there were no mean differences in relational aggression, relational victimization, or drug use at either wave at $p < .004$. Most of the gender differences had small to medium effect sizes (i.e., d$s = .20$ to .40).

Measurement Invariance Across Sites

We also examined measurement invariance across sites. Separate analyses comparing the fit of the six competing models were used to test for configural invariance. These analyses necessitated excluding an item from the drug use scale (i.e., used marijuana) that had a very low base rate that resulted in empty cells for crosstabs of that item with other low frequency items in three of the sites. The seven-factor model again fit the data very well for all four sites (RMSEA = .019 to .020, CFI = .973 to .976, and TLI = .970 to .974) and significantly improved the fit relative to all other models based on the difference test (all ps < .001). Multiple group analyses of the seven-factor model indicated that there were only small decreases in fit for models that imposed scalar invariance across sites (i.e., Model S2, CFI = −.002), and across sites and waves (Model S3, CFI = −.004) (see Table 2).

We also tested for differences in patterns of correlations and means across sites. Constraining factor variances and covariances among factors within each wave across sites resulted in only a slight increase in fit relative to the original model (Model S4, CFI = .003), suggesting that the pattern of correlations among factors was similar across sites. Mean differences across sites were compared by constraining factor means at the Georgia site to zero and determining the extent to which means at each of the other three sites differed from zero. The Georgia site was used as the reference as it was the most different from the other sites (i.e., more rural, higher socioeconomic status, and had the smallest percentage of racial and ethnic minorities). Results of an omnibus Wald test revealed differences in means across sites, $\chi^2(42) = 193.89$, $p < .001$. Follow-up tests of individual means were conducted using a per-test p-value of .001 to maintain a family-wise error rate of $p < .05$. In general the means followed the expected pattern with students from the Georgia site reporting lower means than those at one or more of the other sites for physical

aggression, verbal aggression, delinquent behavior, and drug use. In contrast, there were generally no differences in the reported frequency of relational aggression or of overt or relational victimization.

Analyses of the Seven-Factor Model Based on the Full Sample

After establishing measurement invariance for gender and site, further analyses were conducted on the full sample to test for invariance over time. Scalar invariance was supported based on the small decrease in fit that resulted when thresholds and loadings were constrained to the same values over time (Model F2 versus M1 in Table 2). The 35 unstandardized loadings based on this model ranged from .60 to .92. All but three were .70 or higher (see Appendix Table A1). We next examined the consistency of the correlations among the seven factors over time. Constraining factor variances and within-wave covariances over time resulted in a slight improvement in fit indices (see Model F3). Correlations among the three aggression scales within this model were fairly high (see Table 3). Verbal aggression was highly correlated with physical aggression ($r = .89$) and with relational aggression ($r = .82$). The correlation between physical and relational aggression was also fairly high ($r = .74$).

Results of a Wald test indicated that despite their high intercorrelations, the three aggression factors differed in their pattern of relations with the other four PBFS factors, $\chi^2(8) = 100.83$, $p < .001$. Follow-up tests indicated that all but two of 12 pairwise comparisons were significant at $p < .001$. The overall pattern was consistent with our hypotheses. The Drug Use, Delinquent Behavior and Overt Victimization factors were more highly related to the Physical Aggression factor than to the Relational Aggression factor (differences in rs were .14, .08, and .13, respectively), or the Verbal Aggression factor (differences in rs were .05, .09, and .12, respectively). The Overt Victimization factor was more strongly related to the Verbal Aggression factor than to the Relational Aggression factor (difference in rs = .09). In contrast, the Relational Victimization factor was more strongly related to the Relational Aggression factor than to the Physical Aggression factor (difference in $r = .19$) or the Verbal Aggression factor (difference in $r = .13$).

Relations Between PBFS Factors and Other Concurrent Measures

The final set of analyses examined correlations between the seven PBFS factors and teacher reports of student behavior and student reports on measures of related constructs at Wave 2. These were estimated by incorporating the additional measures into the full sample model of the PBFS that specified scalar invariance. The resulting model fit the data well, $\chi^2(2912) = 9725.84$, RMSEA = .021, CFI = .97, TLI = .96. The concurrent validity of the PBFS factors was supported by their pattern of correlations with teacher ratings of students (see Figure 2). The Delinquent Behavior, Drug Use and Physical Aggression factors were each positively correlated with the BASC Aggression ($r = .20$ to .24) and Conduct Disorder scales ($r = .23$ to .26), and negatively correlated with the Adaptive Behavior composite scale ($r = -.22$ to $-.26$). As expected, they were not significantly correlated with the BASC Anxiety scale and had low correlations (i.e.,.10 or less) with the BASC Depression scale. Correlations between PBFS victimization factors and BASC scales were generally less than .10 in absolute value, with the exception of the correlation between PBFS Relational Victimization and BASC

Depression (r = .17). Differences in the strength of correlations between the PBFS factors and BASC scales were tested using the Mplus estimate function based on p < .001. As hypothesized, the magnitude of correlations with the BASC scales differed across the three PBFS Aggression factors. The BASC Aggression, Conduct Problems, and Adaptive Behavior scales were more strongly related to the PBFS Physical Aggression factor than to the Relational Aggression factor. Their correlations with the PBFS Verbal Aggression factor were generally in between, closer to the magnitude of the correlation with the PBFS Physical Aggression factor in one instance (i.e., with BASC Aggression), and to the PBFS Relational Aggression factor in another (i.e., with BASC Adaptive Behavior). There were no significant differences in correlations between the BASC scales and the two victimization factors.

Correlations between the PBFS factors and student reports on measures of related constructs also showed the hypothesized pattern of relations (see Figure 3). The PBFS Delinquent Behavior, Drug Use, and Aggression factors had moderate to large positive correlations with the Delinquent Peer Associations, Individual Norms and Beliefs About Aggression, and Revenge Goals scales, and moderate negative correlations with the Maintain Relationship Goal scale. There were no significant differences in the correlations of the Delinquent Peer Associations scale with the Delinquent Behavior, Drug Use, and Physical Aggression factors. As would be expected, the two measures of beliefs related to aggression were somewhat more strongly correlated with the Physical Aggression factor than with the Delinquent Behavior and Drug Use factors. There were also differences across the three aggression factors in the strength of their correlations with the other measures. The Delinquent Peer Associations, Norms for Aggression, Beliefs Supporting Aggression, and Maintain Relationship Goal scales were more strongly related to the Physical Aggression factor than to the Verbal Aggression and Relational Aggression factors. In contrast, there were small differences in the patterns of correlations between the three aggression factors and revenge goals. They were, however, much smaller for the two victimization factors than for the other PBFS factors. There were also differences in the patterns of correlations for the two victimization factors. The Delinquent Peer Associations, Individual Norms For Aggression, and Beliefs Supporting Aggression scales were more strongly correlated with the Overt Victimization factor than with the Relational Victimization factor. The Maintain Relationship Goal scale was negatively correlated with the Overt Victimization factor and positively correlated with the Relational Victimization factor.

Discussion

Overall, the results of this study supported the PBFS as a self-report measure of adolescents' frequency of victimization, aggression, and related problem behaviors. The hypothesized seven-factor structure fit the data well, significantly improved the fit relative to several competing models, and demonstrated strong measurement invariance across gender, site and two waves of data separated by over two years. Support was also found for the construct validity of the PBFS. The pattern of differences in factor means was consistent with previous research on gender differences in aggression (see meta-analysis by Card et al., 2008), victimization (e.g., Prinstein, Boergers, & Vernberg., 2001), and other antisocial behaviors (e.g., Moffitt, Caspi, Rutter, & Silva, 2001). The PBFS factors generally showed the

expected pattern of correlations with teacher ratings of adolescents' behavior and with self-report measures of relevant constructs.

There is a long history of both theoretical and empirical support for differentiating between physical and relational aggression and victimization, and our findings are consistent with the broader developmental literature on some key similarities and distinctions in the forms and functions of these constructs. We found that only the Relational Victimization factor was related to depression as measured by the BASC. This finding is consistent with research indicating that compared to physical victimization, relational victimization (Sinclair et al., 2012) and a composite measure of relational and verbal victimization (Cole et al., 2013) were more strongly related to depressive cognitions. Relational versus physical victimization may more directly impact depressive cognitions due to the juxtaposition of its personalized and targeted aim at harming social relationships within a context that is often covert and hard to counter against (Sinclair et al., 2013).

We found fairly clear support for differentiating between physical and relational aggression. Compared with relational aggression, physical aggression was more highly correlated with teacher ratings of aggression and conduct problems, and with adolescent reports of drug use, delinquent behavior and related constructs including delinquent peer associations, and norms and beliefs related to aggression. This is consistent with previous studies that have found stronger relations with delinquency and conduct problems for physical aggression than for relational aggression (Card et al., 2008). Our results are also supported by Moffitt's (1993) theory of adolescent limited delinquency, which emphasizes the role of peer influences on the development of antisocial behavior during adolescence. Our findings also support Cillessen and Mayeux (2004) who suggested that the increased student population in middle as compared to elementary school may result in peer group affiliations among physically aggressive adolescents that reinforce norms and beliefs supporting aggression and the engagement in a variety of externalizing behaviors. These researchers further argued that physical aggression may be driven to a greater extent by individual characteristics whereas relational aggression may be more dependent on contextual factors (e.g., the specific dynamics of social relationships).

Analyses of the PBFS provided fairly clear support for differentiating between physical and relational aggression, but the findings regarding verbal aggression were not as clear. The development of items for the PBFS aggression scales was guided by the assumption that physical, verbal, and relational acts of aggression are best represented by separate, but related factors. The seven-factor model fit the data better than competing models that combined aggression items into one or two factors. Within this model physical and relational aggression were highly correlated (i.e., .74), which is consistent with the average correlation of .76 reported by Card et al. (2008) in their meta-analysis of relations between direct and indirect forms of aggression. Although Verbal Aggression was represented by a separate factor in the seven-factor model, it was highly correlated with both the Physical Aggression ($r = .89$) and Relational Aggression ($r = .82$) factors. Combining verbal aggression items with either physical or relational aggression items resulted in a significant, but fairly small decrease in fit. However, comparison of fit indices for these models did not provide clear support for favoring one model over the other, and combining all three forms of aggression

into a single factor resulted in a clear decrease in fit. There was thus no clear basis for combining verbal aggression with physical aggression versus relational aggression and much less support for combining all three forms into a single measure of aggression.

We found some support for differentiating among physical, relational, and verbal aggression based on differences in their patterns of correlations with other constructs and differences in their means across gender and over time. Teacher ratings on the BASC Aggression scale were more highly correlated with the Physical Aggression and Verbal Aggression factors than with the Relational Aggression factor. This is consistent with the content of the BASC Aggression scale, which includes items representing physical and verbal, but not relational aggression. The pattern of correlations between the Verbal Aggression factor and measures of other constructs was otherwise more similar to the pattern for the Relational Aggression factor. The negative correlation with teacher ratings on the BASC Adaptive Behavior scale was smaller in magnitude for the Verbal Aggression factor than for the Physical Aggression factor. Correlations with adolescent reports of drug use, delinquent behavior, delinquent peer associations, and beliefs supporting aggression were also lower for the Verbal Aggression than for the Physical Aggression factor. Overall, the findings suggest that although both verbal and relational aggression are significantly correlated with other problem behaviors, they represent less extreme forms of problem behavior than physical aggression.

Whereas the literature has been fairly clear in differentiating between physical and relational aggression, it is much less clear where verbal aggression fits within this framework (Ostrov & Kamper, 2015). Although verbal acts of aggression are typically considered a form of overt or direct aggression (Card et al., 2008), Ostrov and Kamper (2015) recently argued against creating composite measures of physical and verbal aggression and called for more research to examine verbal aggression as a distinct construct. The results of the present study highlight the need for further research to determine the value of differentiating among different forms of aggression, particularly verbal aggression. This effort will require more comprehensive measures as many current scales designed to assess overt aggression have only one or two items representing verbal aggression (e.g., Little et al., 2003; Prinstein et al., 2001). The PBFS attempted to address this issue by including a minimum of six items for each form of aggression on the initial version of the scale. However, developing items that unambiguously represent specific forms of aggression can be challenging. This was evident in the analysis of the content of the PBFS items wherein an item originally on the Relational Aggression scale (i.e., "made fun of someone to make others laugh") was moved to the Verbal Aggression scale based on review by a panel of researchers and analysis of part-whole relations with each scale. Further work to evaluate the merits of considering verbal aggression a distinct form of aggression will require appropriate definitions of each form of aggression and development of a pool of items that clearly represents them. Whereas physical and relational aggression have clear distinctions based on the intention to create physical harm versus harm others' social relationships, respectively, such a differentiation is less clear for verbal aggression. Designing items that better clarify the intention of verbal aggression may be helpful in distinguishing this construct from relational and physical aggression or identifying subsets of items that link more specifically to relational or physical aggression. This will provide a basis for further study to determine the value of making distinctions among these forms of aggression.

Although the rationale for differentiating between physical and verbal forms of aggression also applies to victimization, the PBFS Overt Victimization factor did not have an adequate pool of items to create separate factors for each form of victimization. As with aggression, previous studies have differed in their treatment of verbal victimization with some studies finding support for combining it with physical victimization (Rosen, et al., 2013), others incorporating it into relational victimization (Hunt et al., 2012), and still others treating verbal victimization as a distinct factor (Marsh et al., 2011). This suggests the need for further work to examine this issue with more comprehensive measures that address all three forms of victimization. The results of this study supported differentiating between relational and overt forms of victimization. Although the Relational Victimization and Overt Victimization factors were highly correlated ($r = .74$), examination of their pattern of correlations with other variables supported treating them as distinct constructs. As expected, the Physical Aggression, Verbal Aggression, Delinquent Behavior, and Drug Use factors were more highly correlated with the Overt Victimization factor than with the Relational Victimization factor. This is also supported by the stronger correlations found between the Overt Victimization factor and other measures including the BASC Conduct Problem scale and student reports on measures of delinquent peer associations and beliefs related to aggression. This is consistent with prior work demonstrating relations among physical aggression perpetration and victimization and related risk factors (Bettencourt et al., 2013). Further support for discriminant validity is provided by the finding that the Relational Victimization factor was more strongly correlated with Relational Aggression factor than with the Physical Aggression Factor.

This study also provided a strong test of the measurement invariance of the PBFS. Researchers using the same measure for different groups of individuals make an implicit assumption that the underlying structure and properties of the measure will not vary across individuals and over time. Growing recognition of the importance of establishing measurement invariance (e.g., Pitts, West, & Tein, 1996; Widaman & Reise, 1997) has led to increased efforts to examine the consistency of measures of aggression across gender, grade, and over time (e.g., Marsee et al., 2011; Marsh et al., 2011; Rosen et al., 2012). One issue that has received less attention in the literature is the extent to which invariance can be established across samples representing more diverse populations of adolescents. The current study was able to take advantage of a large data set that sampled schools at four sites that differed not only in their location, but in their racial and ethnic composition. Analyses of the PBFS found support for measurement invariance (i.e., item thresholds and loadings) not only across gender and middle school grades, but also across the four sites. This supports the use of the PBFS for assessing aggression, victimization, and problem behaviors for male and female middle school students across grades and across schools serving student populations similar to those examined in the current study.

The results of this study need to be interpreted within the context of the overall pattern of findings and several methodological limitations. Although the hypothesized seven-factor structure fit the data significantly better than the competing models, several competing models fit the data nearly as well. Moreover, although differences were found in the pattern of correlations between the PBFS factors and concurrent measures of related constructs, these differences were often small. This underscores the need for further work to determine

the utility of differentiating among specific forms of aggression. The data from the MVPP provided an opportunity to examine the properties of the PBFS within a large and diverse sample, and to evaluate measurement invariance across schools from different parts of the United States. However, the schools selected for the multi-site study were public schools that served high percentages of students from racial and ethnic minorities, and most were located in urban areas with high rates of crime and poverty (Henry et al., 2004). It is unclear how well these findings might generalize to other samples. Further work is needed to establish measurement invariance of the PBFS across a more diverse range of schools. These data were also collected within the context of an intervention study, which raises the possibility that findings may have been influenced by the intervention. However, analyses indicated strong measurement invariance across measures completed at Wave 1 prior to implementing the intervention and Wave 2, which represented the final post-intervention follow-up assessment.

The PBFS also had several limitations. As previously noted, the pool of items provided a basis for differentiating between verbal aggression and other forms of aggression, but not for differentiating verbal victimization from other forms of victimization. The items were designed to assess the frequency of specific behaviors (e.g., 'put someone down to their face") and thus do not differentiate between types of aggression based on other factors such as the perpetrators' motivation (e.g., Little et al., 2003). The scale may thus be of value in intervention studies or other research focusing on forms of aggression defined by behavior, but of limited value in studies examining other ways of conceptualizing aggression (i.e., proactive or reactive aggression). For future development, the incorporation of items that assess cyber-victimization and aggression will also be important, as will examining how these items fit within the broader structure of the PBFS. Finally, the majority of research on the PBFS has been based on early adolescent samples and additional studies are needed to test its reliability and validity in samples of older adolescents.

Overall, this study supported the PBFS as a self-report measure of adolescents' frequency of victimization, aggression and other problem behaviors. Support was found for its seven-factor structure, which provides scales designed to assess separate forms of both aggression and victimization, and other forms of problem behaviors. The items focus on clearly defined behaviors within a specified period of time (i.e., past 30 days). This is an important feature for interpreting scores of examining changes in the frequency of behavior over time. The PBFS also provides a fairly comprehensive measure that could be useful in evaluations of prevention efforts that target multiple problem behaviors. The current study provided support for measurement invariance of the seven-factor structure across gender, sites, and time. Despite its importance, few prior studies have evaluated the measurement invariance of measures of aggression and victimization. This is a critical property for making meaningful comparisons across groups or over time. This study also provided support for the construct validity of the PBFS. The structure of the PBFS was consistent with theories emphasizing differences across specific forms of aggression. The pattern of differences in factor means was consistent with previous research examining gender differences in rates of aggression (Card et al., 2008), victimization (e.g., Prinstein et al., 2001), and other antisocial behaviors (e.g., Moffitt et al., 2001). Finally, the PBFS factors showed the expected pattern of

correlations with teacher ratings of adolescents' behavior and with other self-report measures of constructs related to aggression and problem behavior.

Acknowledgments

Appendix: PBFS Items, Loadings, and Thresholds

Table A1

Unstandardized loadings and thresholds for the seven-factor measurement model of the Problem Behavior Frequency Scale.

Items	Loadings	Thresholds (SE) 1–2[a]	2–3[b]	3–4[c]
Physical Aggression				
Hit or slapped another kid.	.81 (.01)	0.02 (.02)	0.82 (.02)	1.27 (.02)
Thrown something at another student to hurt them.	.70 (.01)	0.41 (.02)	1.18 (.02)	1.66 (.02)
Threatened to hit or physically harm another kid.	.80 (.01)	0.57 (.02)	1.20 (.02)	1.58 (.02)
Shoved or pushed another kid.	.83 (.01)	−0.28 (.02)	0.64 (.02)	1.09 (.02)
Threatened someone with a weapon (gun, knife, club, etc.).	.77 (.01)	1.58 (.02)	1.96 (.02)	2.19 (.02)
Verbal Aggression				
Put someone down to their face.	.75 (.01)	0.77 (.02)	1.33 (.02)	1.70 (.02)
Picked on someone.	.79 (.01)	−0.20 (.02)	0.57 (.02)	0.96 (.02)
Teased someone to make them angry.	.81 (.01)	0.01 (.02)	0.83 (.02)	1.24 (.02)
Said things about another student make other students laugh.	.80 (.01)	−0.33 (.02)	0.54 (.02)	0.95 (.02)
Relational Aggression				
Told another kid you wouldn't like them unless they did what you wanted them to do.	.72 (.01)	1.10 (.02)	1.77 (.03)	2.09 (.03)
Spread a false rumor about someone.	.78 (.01)	0.77 (.02)	1.43 (.02)	1.77 (.03)
Tried to keep others from liking another kid by saying mean things about him/her.	.76 (.01)	0.65 (.02)	1.35 (.02)	1.69 (.03)
Left another kid out on purpose when it was time to do an activity.	.70 (.01)	0.65 (.02)	1.45 (.02)	1.86 (.03)
Didn't let another student be in your group anymore because you were mad at them.	.60 (.01)	0.16 (.02)	1.08 (.02)	1.60 (.02)
Overt Victimization				
Another student threatened to hit or physically harm you.	.81 (.01)	0.28 (.02)	0.91 (.02)	1.29 (.02)
Been pushed or shoved by another kid.	.84 (.01)	−0.27 (.02)	0.66 (.02)	1.12 (.02)

		Thresholds (SE)		
Items	**Loadings**	**1–2**[a]	**2–3**[b]	**3–4**[c]
Been threatened or injured by someone with a weapon (gun, knife, club, etc.).	.71 (.02)	1.34 (.02)	1.80 (.03)	2.05 (.04)
Been hit by another kid	.83 (.01)	−0.12 (.02)	0.72 (.02)	1.13 (.02)
Been yelled at or called mean names by another kid.	.80 (.01)	−0.27 (.02)	0.55 (.02)	0.93 (.02)
Relational Victimization				
Had a kid who is mad at you try to get back at you by not letting you be in their group anymore.	.79 (.01)	0.41 (.02)	1.09 (.02)	1.46 (.02)
Had a kid say they won't like you unless you do what he/she wanted you to do.	.68 (.01)	0.56 (.02)	1.33 (.02)	1.78 (.03)
Been left out on purpose by other kids when it was time to do an activity.	.72 (.01)	0.40 (.02)	1.14 (.02)	1.53 (.02)
Had someone spread a false rumor about you.	.73 (.01)	−0.13 (.02)	0.77 (.02)	1.29 (.02)
Had a kid try to keep others from liking you by saying mean things about you.	.80 (.01)	0.02 (.02)	0.80 (.02)	1.24 (.02)
Delinquent Behavior				
Stolen something from another student.	.71 (.01)	0.95 (.02)	1.64 (.02)	1.99 (.03)
Snuck into someplace without paying such as movies, onto a bus or subway.	.69 (.01)	1.14 (.02)	1.67 (.02)	1.99 (.03)
Written things or sprayed paint on walls or sidewalks or cars where you were not supposed to.	.77 (.01)	1.36 (.02)	1.84 (.03)	2.11 (.03)
Taken something from a store without paying for it (shoplifted).	.76 (.01)	1.16 (.02)	1.73 (.03)	2.02 (.03)
Damaged school or other property that did not belong to you.	.81 (.01)	1.24 (.02)	1.80 (.02)	2.07 (.02)
Drug Use				
Drunk beer (more than a sip or taste).	.87 (.01)	1.19 (.02)	1.77 (.02)	2.08 (.03)
Drunk wine or wine coolers (more than a sip or taste).	.84 (.01)	1.09 (.02)	1.71 (.02)	2.03 (.03)
Smoked cigarettes.	.84 (.01)	1.48 (.02)	1.93 (.03)	2.15 (.03)
Been drunk.	.89 (.01)	1.65 (.03)	2.06 (.03)	2.35 (.04)
Drunk liquor (like whiskey or gin).	.90 (.01)	1.53 (.02)	2.00 (.03)	2.30 (.04)
Used marijuana (pot, hash, reefer).	.83 (.01)	1.76 (.03)	2.08 (.03)	2.29 (.03)

Note. All loadings significant at $p < .001$. Standard errors in parentheses.

[a.] Threshold between category 1 (i.e., _Never_) and higher frequency categories.

[b.] Threshold between category 2)_1–2 times_) and higher frequency categories.

[c.] Threshold between selecting response option 3 (_3–5 times_) and higher frequency categories.

References

Asparouhov T & Muthén B (2006). Robust Chi Square Difference Testing with mean and variance adjusted test statistics. Mplus Web Notes: No. 10 5 26, 2006.

Barker ED, Tremblay RE, Nagin DS, Vitaro F, & Lacourse E (2006). Development of male proactive and reactive physical aggression during adolescence. Journal of Child Psychology and Psychiatry, 47, 783–790. doi: 10.1111/j.1469-7610.2005.01585.x [PubMed: 16898992]

Bettencourt AF, Farrell AD, Liu W, & Sullivan TN (2013). Stability and change in patterns of peer victimization and aggression during adolescence. Journal of Clinical Child and Adolescent Psychology, 42, 429–441. doi: 10.1080/15374416.2012.738455 [PubMed: 23186101]

Card NA, Stucky BD, Sawalani GM, & Little TD (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and

relations to maladjustment. Child Development, 79, 1185–1229. doi: 10.1111/j. 1467-8624.2008.01184.x [PubMed: 18826521]

Cheung GW, & Rensvold RB (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9, 233–255. doi:10.1207/S15328007SEM0902_5

Cillessen AHN, & Mayeux L (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. Child Development, 75, 147–163. [PubMed: 15015681]

Cole DA, Dukewich TL, Roeder K, Sinclair KR, McMillian J, Will E,…Felton JW (2014). Linking peer victimization to the development of depressive self-schemas in children and adolescents. Journal of Abnormal Child Psychology, 42, 149–160. [PubMed: 23824686]

Crick NR, & Grotpeter JK (1995). Relational aggression, gender, and social-psychological adjustment. Child Development, 66, 710–722. doi: 10.2307/1131945 [PubMed: 7789197]

Crick NR, & Grotpeter JK (1996). Children's treatment by peers: Victims of relational and overt aggression. Development and Psychopathology, 8, 367–380. doi: 10.1017/S0954579400007148

De Los Reyes A, & Kazdin AE (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. Psychological Bulletin, 131(4), 483–509. doi: 10.1037/0033-2909.131.4.483 [PubMed: 16060799]

Desjardins T, Thompson RS, Sukhawathanakul P, Leadbeater BJ, & MacDonald SWS (2013). Factor structure of the Social Experience Questionnaire across time, sex, and grade among early elementary school children. Psychological Assessment, 25, 1058–1068. doi: 10.1037/a0033006 [PubMed: 23730827]

DeVellis RF (2011). Scale development: Theory and applications (3rd ed.). Chapel Hill, NC: Sage Publications.

Embretson SE, & Reise SP (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.

Farrell AD, Ampy LA, & Meyer AL (1998). Identification and assessment of problematic interpersonal situations for urban adolescents. Journal of Clinical Child Psychology, 27(3), 293–305. doi: 10.1207/s15374424jccp2703_6 [PubMed: 9789189]

Farrell AD, & Bruce SE (1997). Impact of exposure to community violence on violent behavior and emotional distress among urban adolescents. Journal of Clinical Child Psychology, 26(1), 2–14. doi: 10.1207/s15374424jccp2703_6 [PubMed: 9118172]

Farrell AD, Henry DB, Mays SA, & Schoeny ME (2011). Parents as moderators of the impact of school norms and peer influences on aggression in middle school students. Child Development, 82(1), 146–161. doi: 10.1111/j.1467-8624.2010.01546.x [PubMed: 21291434]

Farrell AD, Kung EM, White KS, & Valois RF (2000). The structure of self-reported aggression, drug use, and delinquent behaviors during early adolescence. Journal of Clinical Child Psychology, 29, 282–292. doi: 10.1207/S15374424jccp2902_13 [PubMed: 10802836]

Farrell AD, Meyer AL, Sullivan TN, & Kung EM (2003). Evaluation of the Responding in Peaceful and Positive Ways (RIPP) seventh grade violence prevention curriculum. Journal of Child and Family Studies, 12(1), 101–120. doi: 10.1207/s15374424jccp2601_1

Farrell AD, Meyer AL, & White KS (2001). Evaluation of Responding in Peaceful and Positive Ways (RIPP): A school-based prevention program for reducing violence among urban adolescents. Journal of Clinical Child Psychology, 30, 451–463. doi: 10.1207/S15374424JCCP3004_02 [PubMed: 11708233]

Farrell AD, Sullivan TN, Esposito LE, Meyer AL, & Valois RF (2005). A latent growth curve analysis of the structure of aggression, drug use, and delinquent behaviors and their interrelations over time in urban and rural adolescents. Journal of Research on Adolescence, 15(2), 179–204. doi: 10.1111/j.1532-7795.2005.00091.x

Farrington DP (1999). Validity of self-reported delinquency. Criminal Behaviour and Mental Health, 9, 293–295. doi: 10.1002/cbm.327

Furlong MJ, Sharkey JD, Felix ED, Tanigawa D, & Green JG (2010). Bullying assessment: A call for increased precision of self-reporting procedures In Simerson JR, Swearer SM, & Espelage DL (Eds.), Handbook of bullying in schools: An international perspective (pp. 329–346). New York: Routledge.

Galen BR, & Underwood MK (1997). A developmental investigation of social aggression among children. Developmental Psychology, 33, 589–600. doi: 10.1037/0012-1649.33.4.589 [PubMed: 9232374]

Gladden RM, Vivolo-Kantor AM, Hamburger ME, & Lumpkin CD (2014). Bullying surveillance among youths: Uniform definitions for public health and recommended data elements, Version 1.0. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and US Department of Education.

Henry DB, Cartland J, Ruchross H, & Monahan K (2004). A return potential measure of setting norms for aggression. American Journal of Community Psychology, 33, 131–149. doi: 10.1023/B:AJCP. 0000027001.71205.dd [PubMed: 15212174]

Henry DB, Farrell AD, & the Multisite Violence Prevention Project (2004). The study designed by a committee: Design of the Multisite Violence Prevention Project. American Journal of Preventive Medicine, 24, 12–19. doi: 10.1016/j.amepre.2003.09.027

Hopmeyer A, & Asher SR (1997). Children's responses two types of peer conflict situations. Symposium presented at the biennial meeting of the Society for Research in Child Development, Washington, D.C.

Hunt C, Peters L, & Rapee RM (2012). Development of a measure of the experience of being bullied in youth. Psychological Assessment, 24, 156–165. doi: 10.1037/a0025178 [PubMed: 22250594]

Jessor R, & Jessor SL (1977). Problem behavior and psychosocial development: A longitudinal study of youth. New York: Academic Press.

Kandel DB (1975). Stages in adolescent involvement in drug use. Science, 190, 912–914. doi: 10.1126/science.1188374 [PubMed: 1188374]

Kolbe LJ, Kann L, & Collins JL (1993). Overview of the youth risk behavior surveillance system. Public Health Reports, 108(Suppl. 1), 2–10. [PubMed: 8210269]

Little T, Henrich C, Jones S, & Hawley P (2003). Disentangling the "whys" from the "whats" of aggressive behaviour. International Journal of Behavioral Development, 27, 122–133. doi: 10.1080/01650250244000128

Marsee MA, Barry CT, Childs KK, Frick PJ, Kimonis ER, Muñoz LC, … Lau KSL (2011). Assessing the forms and functions of aggression using self-report: Factor structure and invariance of the Peer Conflict Scale in youths. Psychological Assessment, 23, 792–804. doi: 10.1037/a0023369 [PubMed: 21500922]

Marsh HW, Nagengast B, Morin AJS, Parada RH, Craven RG, & Hamilton LR (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. Journal of Educational Psychology, 103, 701–732. doi: 10.1037/a0024122

Mehari KR, Farrell AD, & Le AH (2014). Cyberbullying among adolescents: Measures in search of a construct. Psychology of Violence, 4(4), 1–17. doi: 10.1037/a0037521

Miller-Johnson S, Sullivan TN, Simon TR, and the MVPP (2004). Evaluating the impact of interventions in the Multisite Violence Prevention Study: Samples, procedures, and measures. American Journal of Preventive Medicine, 26, 48–61. doi: 10.1016/j.amepre.2003.09.015 [PubMed: 14732187]

Moffitt TE (1993). Adolescence-Limited and Life-Course-Persistent antisocial behavior: A developmental taxonomy. Psychological Review, 100, 674–701. [PubMed: 8255953]

Moffitt TE, Caspi A, Rutter M, & Silva PA (2001). Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study. doi:.10.17/ CB09780511490057

Orpinas P, & Horne AM (2006). Bullying prevention: Creating a positive school climate and developing social competence. Washington, DC: American Psychological Association. doi: 10.1037/11330-000

Ostrov JM, & Kamper KE (2015). Future directions for research on the development of relational and physical peer victimization. Journal of Clinical Child & Adolescent Psychology, 0(0), 1–11. doi: 10.1080/15374416.2015.1012723

Piquero AR, MacIntosh R, & Hickman M (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. Criminology, 38(3), 897–930. doi: 10.1111/j.1745-9125.2000.tb00910.x

Pitts SC, West SG, & Tein JY (1996), Longitudinal measurement models in evaluation research: Examining stability and change. Evaluation and Program Planning, 19, 333–350. doi:10.1016/S0149-7189(96)00027-4

Prinstein MJ, Boergers J, & Vernberg EM (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. Journal of Clinical Child Psychology, 30, 479–491. doi: 10.1207/S15374424JCCP3004_05 [PubMed: 11708236]

Rescorla LA, Ginzburg S, Achenbach TM, Ivanova MY, Almqvist F, Begovac I,...Verhulst FC (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. Journal of Clinical Child and Adolescent Psychology, 42, 262–273. doi: 10.1080/15374416.2012.717870 [PubMed: 23009025]

Reynolds CR, & Kamphaus RW (1992). Behavior Assessment System for Children: Manual. Circle Pines, MN: American Guidance Service.

Rosen LH, Beron KJ, & Underwood MK (2013). Assessing peer victimization across adolescence: Measurement invariance and developmental change. Psychological Assessment, 25, 1–11. doi: 10.1037/a0028985 [PubMed: 22708574]

Sinclair KR, Cole DA, Dukewich T, Felton J, Weitlauf AS, Maxwell MA,...Jacky A (2012). Impact of physical and relational peer victimization on depressive cognitions in children and adolescents. Journal of Clinical Child & Adolescent Psychology, 41(5), 570–583. doi: 10.1080/15374416.2012.704841 [PubMed: 22867436]

Solberg ME, & Olweus D (2003). Prevalence estimation of school bullying with the Olweus bully/victim questionnaire. Aggressive Behavior, 29, 239–268. doi: 10.1002/ab.10047

Sullivan TN, Farrell AD, & Kliewer W (2006). Peer victimization in early adolescence: Association between physical and relational victimization and drug use, aggression, and delinquent behaviors among urban middle school students. Development and Psychopathology, 18, 119–137. doi: 10.1017/S095457940606007X [PubMed: 16478555]

Thornberry TP, & Krohn MD (2000). The self-report method for measuring delinquency and crime. Criminal Justice, 4, 33–83.

U.S. Department of Health and Human Services. (2001). Youth violence: A report of the Surgeon General. Washington, DC: United States Department of Justice.

Widaman KF, & Reise SP (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain In Bryant KJ, Windle M, & West SG (Eds.), The science of prevention: Methodological advances from alcohol and substance abuse research (pp. 281–324). Washington, DC: American Psychological Association. doi: 10.1037/10222-009
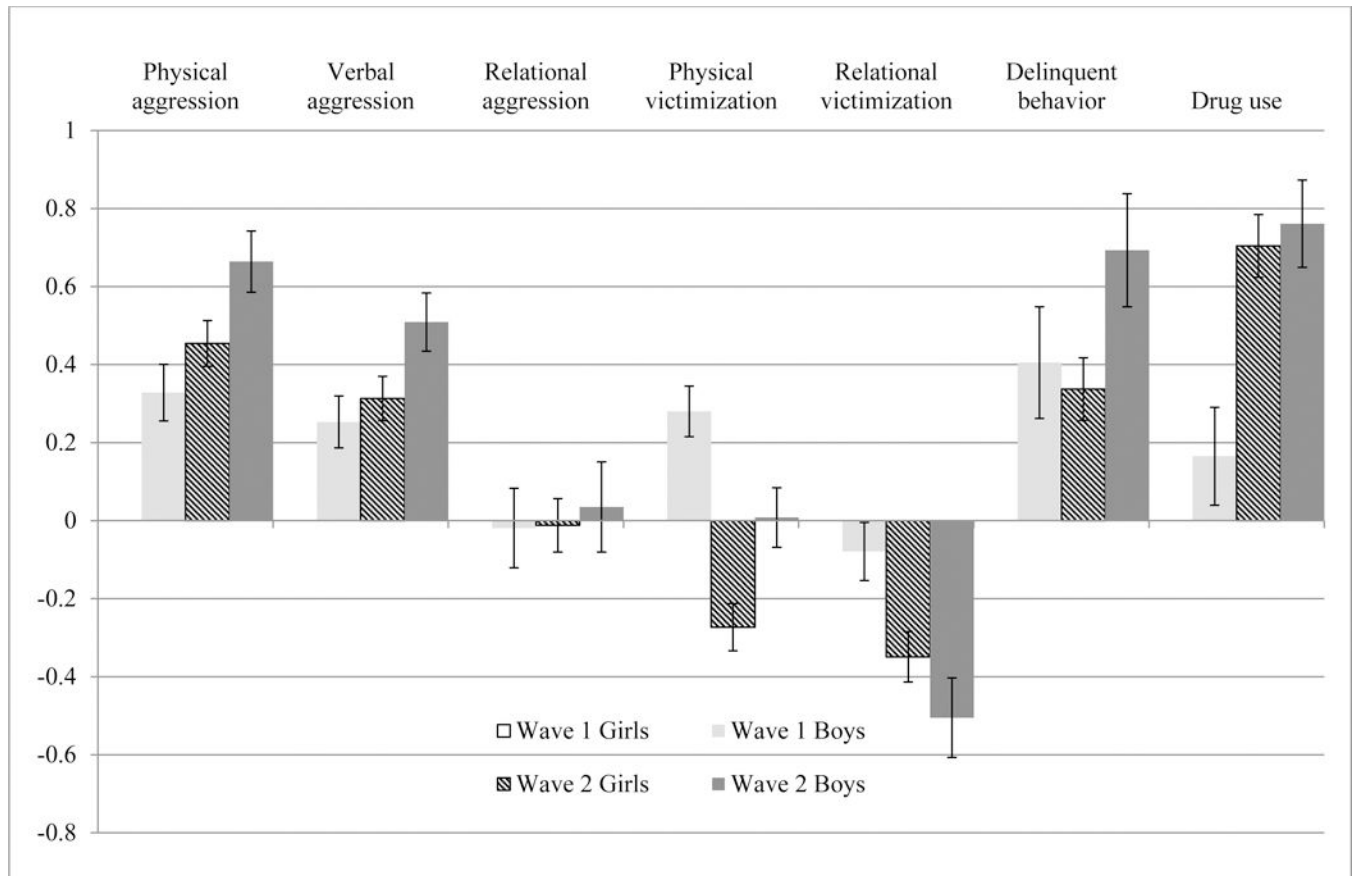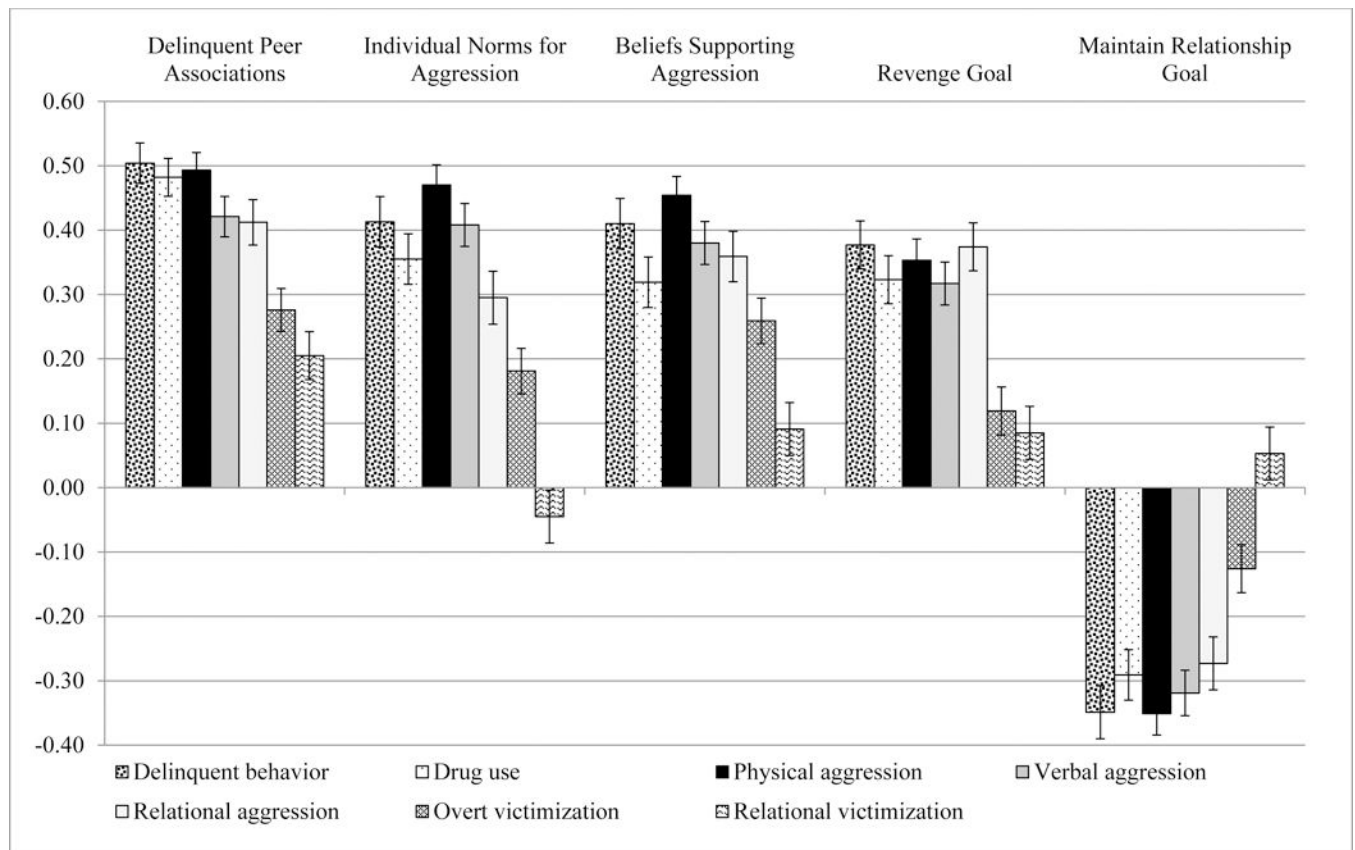
Figure 1:
Confidence intervals (95%) for factor means by gender and wave. Measurement scale for each factor was defined by setting Wave 1 means for girls to zero, and all factor variances were constrained to 1.0.
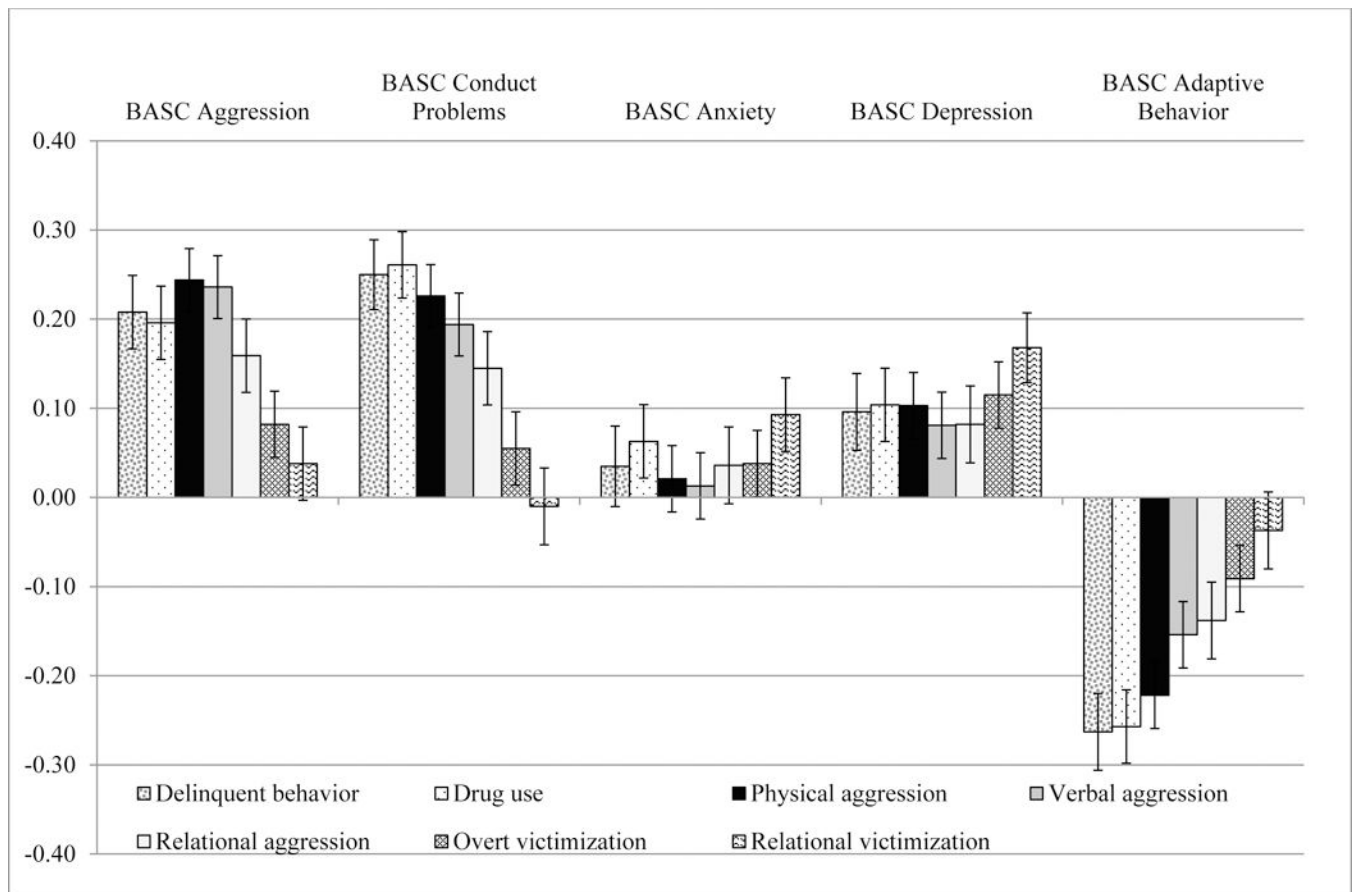
Figure 2.
Confidence intervals (95%) for correlations between PBFS factors and teacher ratings of student behaviors on the Behavioral Assessment System for Children (BASC).

Figure 3.
Confidence intervals (95%) for correlations between PBFS factors and student reports of delinquent peer associations. beliefs related to aggression, and goals for addressing problem situations.

Table 1

Fit indices for competing models of the factor structure of the Problem Behavior Frequency Scale across grades

Model	χ^{2a}	df	$\chi^2\text{diff}^b$	df	RMSEA	CFI	TLI
Competing models of the factor structure							
Seven-factor model (M1)	7872.88*	2219	-	-	.021	.971	.968
Six-factor model combining verbal and physical aggression (M2)	8745.16*	2244	627.91*	25	.023	.966	.964
Six-factor model combined verbal and relational aggression (M3)	9424.03*	2244	838.99*	25	.024	.963	.960
Five-factor overall aggression model (M4)	10936.71*	2265	1620.99*	46	.026	.955	.952
Six-factor overall victimization model (M5)	11719.08*	2244	1410.74*	25	.028	.951	.947
Two-factor problem behavior model (M6)	49955.80*	2304	4586.73*	85	.037	.909	.904
Tests of competing versions of the seven-factor model							
Without serial correlations among measurement errors (M1.1)	8156.07*	2254	606.68*	35	.022	.969	.967
Thresholds invariant across items (M1.2)	51145.42*	2447	36521.33*	228	.060	.748	.751

Note. $N = 5,532$. RMSEA=root mean-square error of approximation. CFI = comparative fit index. TLI = Tucker-Lewis fit index.

[a] Chi-square test of model fit.

[b] Chi-square difference test comparing fit of each model to the seven-factor model such that significant chi-square values indicate that the seven-factor model results in a significant improvement in fit.

*
$p < .001$

Table 2

Fit indices for tests of measurement invariance for the seven-factor model of the Problem Behavior Frequency Scale across gender, site, and time

Model	χ^2	df	RMSEA	CFI	TLI
Multiple Group by Gender					
Configural invariance (G1)	9482.26 *	4438	.020	.973	.970
Scalar invariance across gender (G2)	9729.25 *	4620	.020	.972	.971
Scalar invariance across gender and time (G3)	10183.00 *	4746	.020	.971	.970
Factor variances and covariances constrained across gender (G4)	9291.09 *	4858	.018	.976	.976
Multiple Group by Site [a]					
Configural invariance (S1)	12641.10 *	8340	.019	.975	.972
Scalar invariance across sites (S2)	13472.58 *	8868	.019	.973	.972
Scalar invariance across sites and time (S3)	13933.76 *	8990	.020	.971	.971
Factor variances and covariances constrained across sites (S4)	13376.30 *	9312	.018	.976	.977
Invariance Over Time for the Combined Sample					
Configural invariance (M1)	7872.88 *	2219	-	-	.021
Scalar invariance across time (F2)	8313.80 *	2345	.021	.969	.968
Factor variances and covariances constrained across time (F3)	7235.89 *	2373	.019	.975	.974

Note. $N = 5,532$. χ^2 = chi-square test of model fit. RMSEA=root mean-square error of approximation. CFI = comparative fit index. TLI = Tucker-Lewis fit index.

[a] One drug use item that resulted in estimation problems due to empty cells in one or more group was removed from this analysis.

*
$p < .001$.

Table 3

Correlations among factors within wave (below diagonal) and across waves (on the diagonal).

	PA	VA	RA	OV	RV	DEL	DRG
Physical Aggression (PA)	.50[*]						
Verbal Aggression (VA)	.89[*]	.51[*]					
Relational Aggression (RA)	.74[*]	.82[*]	.46[*]				
Overt Victimization (OV)	.58[*]	.51[*]	.43[*]	.44[*]			
Relational Victimization (RV)	.32[*]	.35[*]	.52[*]	.74[*]	.46[*]		
Delinquent Behavior (DEL)	.81[*]	.73[*]	.74[*]	.45[*]	.30[*]	.48[*]	
Drug Use (DRG)	.66[*]	.55[*]	.55[*]	.30[*]	.20[*]	.80[*]	.49[*]
Wave 2 Means[a]	.38[*]	.28[*]	−.02	−.28[*]	−.43[*]	.29[*]	.61[*]

Note. $N = 5,532$. Estimates based on seven-factor model with loadings and thresholds constrained across waves. All factor variances were constrained to 1, and intercorrelations among factors within each wave were constrained to the same values across waves.

[a] Wave 1 means were constrained to zero to make the model identifiable.

[*] $p < .001$.