10-1-2018

Evaluation of the Problem Behavior Frequency Scale-teacher Report Form for Assessing Behavior in a Sample of Urban Adolescents

Albert D. Farrell
Virginia Commonwealth University, afarrell@vcu.edu

Elizabeth A. Goncy
Virginia Commonwealth University, e.goncy@csuohio.edu

Terri N. Sullivan
Virginia Commonwealth University

Erin L. Thompson
Virginia Commonwealth University

HHS Public Access
Author manuscript
Psychol Assess. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:
Psychol Assess. 2018 October ; 30(10): 1277–1291. doi:10.1037/pas0000571.

Evaluation of the Problem Behavior Frequency Scale – Teacher Report Form for Assessing Behavior in a Sample of Urban Adolescents

Albert D. Farrell, **Elizabeth A. Goncy**[1], **Terri N. Sullivan**, **Erin L. Thompson**
Virginia Commonwealth University

Abstract

This study evaluated the structure and validity of the Problem Behavior Frequency Scale – Teacher Report Form (PBFS-TR) for assessing students' frequency of specific forms of aggression and victimization, and positive behavior. Analyses were conducted on two waves of data from 727 students from two urban middle schools (Sample 1) who were rated by their teachers on the PBFS-TR and the Social Skills Improvement System (SSIS), and on data collected from 1,740 students from three urban middle schools (Sample 2) for whom data on both the teacher and student report version of the PBFS were obtained. Confirmatory factor analyses supported first-order factors representing three forms of aggression (physical, verbal, and relational), three forms of victimization (physical, verbal and relational), and two forms of positive behavior (prosocial behavior and effective nonviolent behavior), and higher-order factors representing aggression, victimization, and positive behavior. Strong measurement invariance was established over gender, grade, intervention condition, and time. Support for convergent validity was found based on correlations between corresponding scales on the PBFS-TR and teacher ratings on the SSIS in Sample 1. Significant correlations were also found between teacher ratings on the PBFS-TR and student ratings of their behavior on the Problem Behavior Frequency Scale–Adolescent Report (PBFS-AR) and a measure of nonviolent behavioral intentions in Sample 2. Overall the findings provided support for the PBFS-TR, and suggested that teachers can provide useful data on students' aggressive and prosocial behavior and victimization experiences within the school setting.

Keywords

teacher ratings; assessment of aggression; assessment of victimization; assessment of problem behaviors in adolescence; measurement invariance

Research to identify the causes and consequences of aggression and victimization, and efforts to develop and evaluate school-based violence prevention efforts require well-developed measures of adolescents' behavior. Researchers have used a variety of methods to assess aggression and victimization. These include youth self-report, ratings of youth by

Correspondence concerning this article should be addressed to Albert D. Farrell, Department of Psychology, Virginia Commonwealth University, P.O. Box 842018, Richmond, VA 23284-2018. afarrell@vcu.edu.
[1]Now at Cleveland State University

parents and teachers, archival data (e.g., school office discipline referrals), and observations (e.g., Dahlberg, Toal, Swahn, & Behrens, 2005). Studies have generally found low to moderate agreement across measures of these constructs obtained from different informants (e.g., Farrell, Sullivan, Goncy & Le, 2016). No one method has yet to emerge as the single best approach, nor is that likely because each has its own inherent strengths and limitations. This may reflect different sources of bias (e.g., self-report versus rater biases) and differences related to the context of observation. For example, parents and teachers each observe behavior in different settings. This suggests the need for well-developed measures from multiple sources to provide a complete picture of behavior and to determine the extent to which findings vary as a function of the source of data (De Los Reyes & Kazdin, 2005). Although a host of self-report measures of aggression and victimization have been developed (Furlong, Sharkey, Felix, Tanigawa, & Green, 2010), fewer efforts have been made to develop measures based on teacher report. The goal of this study was to evaluate the teacher report form of the Problem Behavior Frequency Scale (PBFS-TR) – a measure designed to provide data on students' aggressive and prosocial behavior and victimization experiences within the school setting.

Teachers provide a particularly relevant source of information regarding adolescents' behavior. They often spend more time during the day with adolescents than do parents and they observe them in both structured (e.g., classroom) and unstructured (e.g., lunchroom) settings with their peers. Teachers also have experience with different students from the same age group over extended periods of time, enabling them to develop informal norms for evaluating student behavior more objectively than parents are able to do. As a result, teachers are often the first to identify behavioral problems (Orpinas, Raczynski, Peters, Colman, & Bandalos, 2015). Teacher ratings may have particular value for school-based research projects. The fact that they are limited to observations of behavior at school make them especially relevant for evaluating school-based interventions. Teacher ratings combined with data from other sources that assess behavior outside of school may also provide a basis for determining the extent to which behavior varies across contexts. At a practical level, it may be less costly to collect data from teachers than from parents of individual students (Clemans, Musci, Leoutsakos, & Ialongo, 2014).

Teacher-report measures are often limited in the information they provide about students' aggressive behavior. Researchers have emphasized the importance of differentiating among types of aggression that differ in their form. These include direct forms of aggression such as physical and verbal acts, and indirect forms such as acts of relational or social aggression that are designed to damage social relationships. Reviews of the literature have highlighted the importance of these distinctions in terms of differences in their prevalence, causes, and consequences for youth (e.g., Card, Stucky, Sawalani, & Little, 2008). Some teacher report measures include aggression or bullying scales, but do not differentiate among forms of aggression (e.g., Gresham & Elliott, 2008; Reynolds & Kamphaus, 2004). Others assess only a single form of aggression (e.g., Vitaro et al., 2016). Assessing multiple forms of aggression can provide useful information. For example, a recent evaluation of a school-based violence prevention program found that effects on verbal and relational aggression emerged during the second year of implementing an intervention, but were not evident for physical aggression until the third year of implementation (Author reference). Such sequenced effects would not be evident with more global measures.

Although studies examining adolescents' self-report measures have found support for distinct factors representing different forms of aggression (e.g., Card et al., 2008), few studies have investigated whether teachers can differentiate among specific forms of aggression. Whereas physical aggression can be readily observed, other forms such as relational aggression are more subtle and require an informed understanding of peer group structures. The Children's Social Behavior Questionnaire – Teacher Version (Crick, 1996) is one of the few teacher-report measures of aggression that differentiates between overt and relational aggression. Crick (1996) found that teacher-reported relational aggression uniquely contributed to predicting future peer rejection while controlling for physical aggression, underscoring the importance of examining multiple forms. Although widely used, empirical support for its psychometric properties is limited to exploratory analyses focused on elementary school children and evaluations of the internal consistency of individual scales. Moreover, there has been little to no research examining whether teacher reports of verbal acts of aggression should be considered distinct from physical and relational aggression.

Teacher rating scales are also often limited in their assessment of student's victimization experiences, particularly the ability to differentiate among forms of victimization. The Social Experience Questionnaire -Teacher Report (Cullerton-Sen & Crick, 2005) is one of the few teacher-report measures that assesses both relational and physical victimization. However, it has only three items for each form of victimization, does not examine verbal victimization separately, and its structure has not been empirically verified using confirmatory analyses. It thus remains an open question as to whether teacher-rating scales can provide useful information about the specific forms of victimization experienced by adolescents.

There is value in assessing not only problem behaviors, but also positive behaviors. Prosocial behavior is a core dimension of adolescent social competence (Gresham, Cook, Crews, & Kern, 2004; Wentzel, Filisetti, & Looney, 2007). Effective nonviolent behavior is a related construct, defined as a response to a problematic situation that maximize positive consequences while minimizing negative ones (Goldfried & D'Zurilla, 1969). This definition includes two parts: the response is nonviolent and it effectively addresses the problem. For example, avoiding a friend with whom you are having a disagreement may be nonviolent, but is not effective. In contrast, attempting to talk it out could be both nonviolent and effective. Research has shown that nonaggressive youth represent a heterogeneous group of individuals, not all of whom engage in effective nonviolent and prosocial behavior (Farrell et al., 2007). This highlights the need to identify factors that promote adolescents' use of responses to problem situations that are both nonviolent and effective. This could also inform school-based violence prevention efforts that have been shown to be most effective when concurrent efforts are made not only to reduce problem behavior but also to support nonviolent and prosocial behavior (for a review see Greenberg et al., 2003).

The Problem Behavior Frequency Scale – Teacher Report (PBFS-TR)

The PBFS-TR was developed to assess multiple forms of aggression and victimization and two forms of positive youth behavior (i.e., prosocial behavior, effective non-violent behavior) among middle school students. It has several features that distinguish it from many other teacher rating scales. It was designed to assess multiple forms of aggression and victimization using items based on the adolescent-report version of the PBFS (PBFS-AR; Farrell et al., 2016). The PBFS-TR also includes items representing effective nonviolent behavior derived from mixed-methods studies in which adolescents, parents, and community representatives rated the effectiveness of specific responses to problem situations (Author references). Although evaluation of the adolescent report version of the PBFS found support for three forms of aggression (physical, verbal, and relational) and two forms of victimization (overt and relational) (Farrell et al., 2016), it is unclear whether teachers can also differentiate among these same forms of aggression or reliably report on students' victimization.

The Present Study

The purpose of this study was to evaluate the structure and validity of the PBFS-TR for assessing adolescents' behavior. We hypothesized that support would be found for three aggression factors (Physical, Verbal, and Relational), three victimization factors (Physical, Verbal, and Relational), and two positive behavior factors (Prosocial and Effective Nonviolent). We also evaluated measurement invariance. Measurement invariance is a critical factor in determining the extent to which measurement properties can be generalized across individuals and contexts (Widaman & Reise, 1997). A critical, but typically untested assumption in using a measure to evaluate the impact of an intervention is that the intervention will not influence the measure's structure or measurement properties. However, implementing a school-level violence prevention program could sensitize teachers to different forms of aggression, which could alter the structure of a measure based on teacher ratings. This could complicate comparing scores across intervention conditions. This project involved secondary analysis of data collected from two studies in which teacher ratings were obtained to evaluate the impact of school-based violence prevention programs. This provided an opportunity to evaluate measurement invariance across individuals in intervention and control conditions. In addition, we examined measurement invariance across gender, grade and over time.

We also investigated the convergent validity of scores on the PBFS-TR based on its correlations with teacher ratings on the Social Skills Improvement System Rating Scales (SSIS; Gresham & Elliott, 2008). We hypothesized that the PBFS-TR aggression scales would be positively correlated with the broader SSIS measure of bullying, and negatively correlated with SSIS Self-Control, Responsibility, and Empathy scales. Conversely, we hypothesized that the PBFS-TR Prosocial and Effective Nonviolent Behavior scales would be positively correlated with SSIS Self-Control, Responsibility, Empathy, and Academic Competence scales, and negatively correlated with the Bullying scale. The PBFS-TR Victimization factor did not have a counterpart on the SSIS. Given the strong relation between aggression and victimization, we hypothesized that its pattern of relations with

SSIS scales would be similar to the pattern expected for aggression, but with lower correlations. Based on the relation between victimization and internalizing problems (e.g., Reijntjes, Kamphuis, Prinzie, & Telch, 2010), we further hypothesized that the PBFS-TR Victimization scale would be positively correlated with the SSIS Internalizing Problems scale.

Finally, we examined relations between the teacher and student report versions of the PBFS. We hypothesized that the Aggression and Victimization factors would be positively correlated across informants. We hypothesized that the PBFS-TR Aggression factor would be positively correlated with PBFS-AR factors representing other problem behaviors (e.g., substance use and delinquent behavior). Because the PBFS-AR does not assess positive behaviors we examined relations between the PBFS-TR Prosocial and Effective Nonviolent Behavior factors and a student-report measure of intentions for nonviolent behavior. We hypothesized that PBFS-TR Prosocial and Effective Nonviolent Behavior factors would be positively correlated with intentions to use nonviolent responses in problem situations, and negatively correlated with PBFS-AR Problem Behavior and Victimization factors.

Method

This project was based on secondary analysis of data from two independent samples of students from five public middle schools in an urban school system in the southeastern United States. The schools served a predominantly African American student population most of whom (i.e., over 96%) were eligible for the federal free or reduced lunch program. All procedures were approved by the University's institutional review board.

Sample 1

Students were recruited from classrooms in two middle schools as part of a randomized trial (Author reference) to evaluate the Second Step intervention (Committee for Children, 2008). The project was conducted at one school during the 2011–2012 school year and at a second school during the 2012–2013 school year. Parental consent, student assent, and teacher ratings were obtained for 732 students, which represented 71% of all eligible students. The majority (94%) described themselves as Black or African American including 16% who also endorsed another race; 9% described themselves as Hispanic or Latino. The sample was 52% female and fairly evenly divided across grades 6, 7, and 8 (*Ns* = 230 to 254). Ages ranged from 11 to 16 (median = 12). In terms of family structure, 20% were in two-parent families, 42% were in single-mother households including 13% that included another adult, and 6% lived with a relative and neither parent.

Classrooms within each school were randomly assigned to intervention or control conditions. Teacher ratings on the PBFS-TR and SSIS were obtained for 727 students (99.6% of those consented) near the beginning of the school year (Wave 1), and for 660 (90.4%) of these students near the end of the school year (Wave 2). All Wave 1 data were collected at pretest prior to implementing the intervention. Wave 2 data were obtained at posttest such that 40% of the students were in classrooms where the Second Step curriculum had been implemented.

Sample 2

Data were collected for 1,740 students at three middle schools as part of a study designed to evaluate school-wide implementation of the Olweus Bullying Prevention Program (e.g., Olweus & Limber, 2010). Students were recruited from a random sample of sixth, seventh, and eighth grade students from the school rosters during the project's first year. After the first year, a new sample of entering sixth graders was recruited and seventh and eighth graders were recruited to replace those who left the study. Students continued to participate each year until they left the school or chose to discontinue. Parental consent, student assent, and teacher ratings were obtained from 80% of those eligible. The sample was 53% female and had 579 to 582 students from each grade. Most (i.e., 90%) described themselves as Black or African American; 15% as Hispanic or Latino. Ages ranged from 10 to 16 (median = 12). The most frequently reported family structures were living with a single mother (40%), both biological parents (26%), and a relative without either parent (6%).

The project providing the data used a school-level multiple baseline design in which the intervention was implemented in one school beginning in the second year of data collection, in a second school beginning in the third year, and was not implemented at the third school during the five years of data collection (Author reference). The project used randomization to determine the order and timing of implementation in the three schools. Teachers completed ratings of students and students completed self-report measures during the beginning, middle, and end of school years between 2010 and 2015. The project used a planned missingness design in which students were randomly assigned to be assessed at one or two waves during the school year. Although most students participated at more than one wave, we focused the current study on a cross-sectional dataset constructed by randomly sampling one wave from each participant using a strategy that resulted in an even distribution across grades and time of year. This provided a basis for between-subject comparisons to examine differences in the factor structure across grades and intervention conditions. About half (i.e., 48%) of the cases in the resulting dataset were for students who were at one of the schools during a time when the intervention was not being implemented.

Measures

For both samples, a core academic teacher completed ratings of each student on the PBFS-TR. For Sample 1, 56 teachers rated between 1 and 24 students ($M = 13$) at each wave. These same teachers also completed the SSIS. For Sample 2, 151 teachers rated between 1 and 44 students ($M = 12$). Students in Sample 2 completed the PBFS-AR and Behavioral Intentions measures using a computer-assisted interview in which they could see each item presented on a laptop computer screen and hear it read through an audio recording. These measures were completed at the same wave when teachers completed the PBFS-TR.

Problem Behavior Frequency Scale – Teacher Report Form (PBFS-TR).—The PBFS-TR is a teacher report form of the PBFS-AR (Farrell et al., 2016). The version administered to Sample 1 included 42 items with subscales designed to assess physical, verbal, and relational aggression, physical and relational victimization, prosocial behavior, and effective nonviolent behavior. The version administered to Sample 2 included an additional verbal aggression item, and three items to assess verbal victimization (e.g., see supplemental materials for a full copy of the measure). Teachers rate how frequently the identified adolescent engaged in or experienced each behavior in the past 30 days using a 4-point scale, where 1 = *Never,* 2 = *Sometimes,* 3 = *Often* and 4 = *Very Often.*

The Social Skills Improvement System – Teacher Form (SSIS; Gresham & Elliott, 2008) (Sample 1 only).—The SSIS is a widely-used, nationally normed measure that assesses social skills and problem behaviors. For this study, we focused on the Self-Control, Responsibility, Empathy, Academic Competence, Internalizing Problems, and Bullying scales. Each of these scales is based on five to seven items for a total of 37 items. Teachers rate the frequency of each item in the past 2 months on a 4-point scale, where 1 = *Never,* 2 = *Seldom,* 3 = *Often* and 4 = *Almost Always.* Previous analyses have found moderate to high intercorrelations and item-total correlations across forms by age (Gresham & Elliot, 2008). Moderate to high correlations have also been found between the SSIS and the Behavioral Assessment System for Children, Second Edition subscales (BASC-2; Reynolds & Kamphaus, 2004). Within the current study, the internal consistencies for all six scales were all above .90 at both waves (alphas = .90 to .98), except for Internalizing Problems which had alphas of .82 and .83 at waves 1 and 2, respectively.

The Problem Behavior Frequency Scale – Adolescent Report (PBFS-AR; Farrell et al., 2016) (Sample 2 only).—The PBFS-AR was designed to assess the frequency of physical, verbal, and relational forms of both aggression and victimization, as well as substance use and other delinquent behaviors. Each scale consists of four to eight items (total of 37) that are rated on a 6-point frequency scale based on the past 30 days, 1 = *Never,* 2 = 1–2 *times,* 3 = 3–5 *times,* 4 = *6–9 times,* 5 = *10–19 times,* and 6 = *20 or more times.* Farrell et al. (2016) found support for separate factors representing physical aggression, verbal aggression, relational aggression, substance use, and other delinquent behavior, and overt and relational victimization, and strong measurement invariance across gender, four geographic locations, and middle school grades. They also established the convergent validity of scores based on their correlations with teacher ratings on the Behavioral Assessment System for Children (Reynolds & Kamphaus, 2004) and self-report measures of relevant constructs. The current project used an updated version of the PBFS-AR based on a more recent evaluation of the structure of the measure in a large sample of urban adolescents (Author reference). We examined the relation between the PBFS-TR and the following five factors on the updated version: Physical Aggression, Relational Aggression, Substance Use, Delinquent Behavior, and Victimization (including physical, verbal, and relational victimization).

Intentions for Nonviolent Behavior (Sample 2 only).—The Intentions for Nonviolent Behavior subscale is a subscale of the Behavioral Intentions scale (Author reference). The Behavioral Intentions scale is based on peer conflict situations and potential ways of responding to them identified in qualitative studies conducted with predominantly African American samples of youth from urban middle schools (Author references). The full measure presents nine hypothetical peer conflict situations that are each followed by two possible responses. Responses include specific physically and relationally aggressive responses, ineffective nonviolent responses, and effective nonviolent responses. Responses were identified as effective or ineffective based on ratings by youth and representatives from the community. Adolescents rate how likely they are to make each response in that situation using a 5-point scale: 1 = *Definitely would not,* 2 = *Probably would not,* 3 = *Might or might not,* 4 = *Probably would,* and 5 = *Definitely would.* The Intentions for Nonviolent Behavior scale is based on ratings of the five effective nonviolent responses (e.g., apologizing to a friend, declining to fight, and calmly discussing a conflict with a friend). This measure was not administered during the last year of the project because of a need to reduce the time required for students to complete the full set of measures. As a result, scores on this measure were available for only 1,301 (i.e., 75%) of the participants in the dataset for the current study.

Analysis

We conducted all analyses using Mplus 7.11. We first ran confirmatory factor analyses for each domain (i.e., aggression, victimization, and positive behaviors) to compare competing models for the structure of aggression, victimization, and positive behavior. We analyzed the two waves of Sample 1 using longitudinal models in which we allowed parameter estimates to vary across waves. The models included serial correlations for measurement errors across waves. We conducted cross-sectional analyses of Sample 2 using the single wave of data included for each adolescent in that sample. In each case, we compared the hypothesized factor structure to alternative models that specified a smaller number of factors.

All models took the clustering of students within teachers into account using the Mplus type=complex option. This approach uses a sandwich estimator to compute the standard errors and chi-square test (Muthén & Satorra, 1995). We treated items as ordered categorical variables using weighted least squares mean- and variance-adjusted estimators (WLSMV). This is comparable to a graded response item-response theory model (Embretson & Reise, 2000). We compared the fit of competing models based on the difference test for WLSMV calculated by Mplus (Muthén & Asparouhov, 2006), and model fit based on the root mean square error of approximation (RMSEA), Tucker-Lewis fit index (TLI), and comparative fit index (CFI). We followed Hu and Bentler’s (1999) recommendation and considered models to have a good fit based on cutoffs of close to .95 or higher for the CFI and TLI, and close to .06 or lower for the RMSEA. The WLSMV estimator uses a pairwise-present approach to addressing missing data (Muthén & Muthén, 2015, p. 8). In Sample 1 the average amount of missing data across SSIS scales was 1.8% and 1.2% for participants at waves 1 and 2, respectively, and 1.2% and 1.1% across PBFS-TR items for participants at waves 1 and 2, respectively. For Sample 2 the average amount of missing data across items was 2.8% for the PBFS-TR, and 2.5% for the PBFS-AR, and was higher (i.e., 29%) for the Intentions for Nonviolent Behavior scale, which was not administered during the final year of the project.

After establishing the overall structure, we used the data from Sample 1 to test for measurement invariance over intervention conditions through multiple group analysis of the Wave 2 data, and examined invariance over time using longitudinal data from the full sample. We conducted multiple group analyses of the cross-sectional data from Sample 2 to evaluate measurement invariance over gender, grade, and intervention condition. In each case we first tested configural invariance based on models that specified the same structure (i.e., specification of which items load on which factor) over time or across groups but that allowed parameter estimates to differ. We then evaluated strong measurement invariance by comparing the fit of the configural invariance models to models that constrained the factor loadings and item thresholds to the same values over time or across groups. Although analyses that treat items as continuous often test an intermediate step of metric invariance, this is not done in analyses of categorical indicators, which focus on the distribution of item categories (Muthén & Muthén, 2015). Because of the large sample size and resulting power to detect minor differences in fit, we followed recommendations by Cheung and Rensvold (2002) to consider measurement invariance satisfied if imposing strong measurement invariance did not decrease the CFI by .01 or more. Finally, we evaluated the validity of the PBFS-TR for assessing adolescents' behavior by examining correlations between PBFS-TR scales and scores on the SSIS in Sample 1, and the PBFS-AR and Intentions for Nonviolent Behavior scale in Sample 2.

Results

Preliminary Analyses

Although teachers rated PBFS-TR items on a 4-point scale, the highest category *(very often)* was rarely endorsed (2% or less) for the victimization and relational aggression items. Such low frequencies create estimation problems for the WLSMV estimator. Inspection of item characteristic curves based on initial models suggested little differentiation between the two highest categories (i.e., *often* and *very often)* for these items. We therefore combined them for these items. We also recoded the item "Been in a fight in which someone was hit" into two categories (Never versus one or more occurrence) based on the low frequency of endorsement for the two highest categories. Item information curves for three other items (threatened someone with a weapon, threatened a teacher, been threatened or injured by someone with a weapon) that had extremely low base rates (see supplemental Table S-1) indicated that they contributed minimal information to the overall reliability of their hypothesized factors. Because their restricted range also created estimation problems we excluded these items from subsequent analyses.

Structure of Aggression

The hypothesized three-factor model for aggression with separate factors for physical, verbal, and relational aggression fit the data very well in Sample 1 (see Model 1 in Table 1) and Sample 2 (see Model 1 in Table 2). Both RMSEAs were less than .05, and CFIs and TLIs were greater than .98. Within this model, the Verbal Aggression factor was highly correlated with the Physical Aggression (*rs* = .88 to .91) and Relational Aggression (*rs* = .86 to .90) factors. The correlation between the Physical Aggression and Relational Aggression factors was also quite high (*rs* = .80 to .82). The difference test indicated that the three-factor model fit significantly better than competing models that specified a smaller number of factors. However, the CFI for the three-factor model was only slightly better than two-factor models that combined verbal with physical aggression into an Overt Victimization factor, or verbal with relational aggression into a Nonphysical Aggression factor. The three-factor model did, however, clearly improve the fit compared with the model that specified a single overall factor (see tables 1 and 2). Though these findings suggested that combining verbal aggression items with either relational or physical aggression items fit the data nearly as well as a model that differentiated among the three forms of aggression, there was little basis for choosing one of these two-factor models over the other. Moreover, there was not support for representing all of the aggression items by a single factor. Based on these results, we chose to conduct further analyses based on the three-factor model of aggression.

Structure of Victimization

The two-factor model with separate factors representing physical and relational victimization fit the data well in Sample 1 and resulted in a significant improvement in fit compared with a one-factor model (see Table 1). Within this model the Physical Victimization and Relational Victimization factors were highly correlated (rs = .79 and .70 for waves 1 and 2, respectively). Participants in Sample 2 completed a version of the PBFS-TR containing additional items that enabled us to test a three-factor model with separate factors representing physical, verbal, and relational victimization. This model fit the data very well (see Table 2), and significantly improved upon the fit of the two-factor and one-factor models based on the difference test and increase in the CFI. Within this model, the Verbal Victimization factor was highly correlated with the Physical Victimization and Relational Victimization factors (both rs = .85). The correlation between Physical Victimization and Relational Victimization factors was lower (r = .76).

Structure of Positive Behaviors

The two-factor model of positive behaviors with separate Prosocial and Effective Non-Violent Behavior factors fit significantly better than the one-factor model in both samples. However, the CFIs and TLIs were low (.88 to .94) and the RMSEA was above .08 in Sample 1. We therefore conducted an exploratory factor analysis using data from Sample 2 to identify plausible alternative models. The results of these analysis supported separate factors for Prosocial and Effective Nonviolent Behavior, but identified two items from the Effective Nonviolent Behavior scale that loaded on a third factor. Both items involved seeking help from an adult. These items differed from the other effective nonviolent behavior items that involved taking direct action to address the situation (see Supplemental Table S1). Excluding these two items improved the fit of the two-factor model not only in Sample 2 on which the exploratory analysis was conducted, but also in Sample 1 (see tables 1 and 2). This revised two-factor model fit significantly better than the one-factor model at $p < .001$, and increased the CFI by .024 in Sample 1 and by .030 in Sample 2. Within this model, the correlation between the two factors ranged from .74 to .84.

Overall Structure of the PBFS-TR

We next examined the overall structure of the PBFS-TR by combining the submodels for each domain into a single model with separate factors representing each form of aggression, victimization, and positive behavior. The resulting seven-factor model for Sample 1 fit the data very well (see Model 11 in Table 1). As in the submodels, there were high correlations among the three aggression factors (rs = .80 to .91), and between the two positive behavior factors (r = .73 and .83 at waves 1 and 2, respectively). There was also a high correlation between the Physical Aggression and Physical Victimization factors (rs = .80 and .85 at waves 1 and 2, respectively). The overall model for Sample 2 included the additional Verbal

Victimization factor. Although this eight-factor model fit the data very well (see Model 13 in Table 2), linear dependencies among the eight latent variables resulted in estimation problems.

The high correlations among the factors and failure to obtain a proper solution in Sample 2 led us to conduct further analyses using exploratory structural equation modeling (Asparouhov & Muthen, 2009). Marsh et al. (2009) observed that the typical model used in confirmatory factor analysis, which requires that each item load on a single factor, can result in poorly fitting models that distort relations among the resulting factors. More specifically, they argued that the exclusion of significant non-zero cross-loadings can overestimate correlations among factors. Although the models we tested fit the data very well, we considered the possibility that some items may have represented multiple factors. We investigated this by testing an exploratory structural equation model (Asparouhov & Muthen, 2009; Marsh et al., 2009) that allowed each item to load on all of the factors using a target rotation based on the hypothesized factor structure. Our goal was to identify items for which we could justify cross-loadings based on both empirical findings and substantive considerations.

We conducted an analysis that included all of the aggression and victimization items, and a separate analysis of the positive behavior items. We began by identifying items with significant cross-loadings exceeding .22 (i.e., 5% shared variance) in both waves of Sample 1 and in Sample 2. Although this is below the typical cutoff of .33 (i.e., 10% shared variance), we required that it be replicated in all three samples to reduce the likelihood of spurious sample-specific findings. Five items in our analysis of the aggression and victimization items, and one item in our analysis of the positive behavior items met this criterion. We next reviewed this pool of items to determine if a cross-loading was justified based on substantive considerations. Two items from the relational aggression scale had cross-loadings on verbal aggression, but were clear examples of indirect acts of relational aggression (i.e., "spread a false rumor about someone" and "tried to keep others form liking another kid"). Because these were not consistent with direct verbal aggression, we did not consider it appropriate to include cross-loadings on the Verbal Aggression factor. In contrast, there were three items meeting our empirical criteria where including cross-loadings seemed justified. Two were from the Physical Aggression factor. The item, "threatened to hit or physically harm someone" was also related to the Verbal Aggression factor, and the item, "was in a fight in which someone was hit" was also related to the Physical Victimization factor. A third item, which was from the Effective Nonviolent Behavior factor, "Apologized to someone when she or he was wrong," was also related to the Prosocial Behavior factor.

Cross-loadings typically indicate that an item does not adequately represent the construct it was designed to assess. In some instances, researchers may opt to delete such items from a measure. In this case we believe that each of these three items may be considered legitimate indicators of more than one factor. The item "threatened to hit or physically harm someone" involves both a verbal component, but also includes a threat of physical harm that goes beyond other acts of verbal aggression such as insults and teasing. Unlike other physical aggression items that indicate a clear direction (e.g., "hit or slapped someone"), teachers who observe a student in a physical fight may not have a basis for identifying the aggressor and the victim. The fact that it loaded on both physical aggression and physical victimization is therefore not surprising. The item “apologized to someone when she or he was wrong” was originally intended to reflect a nonviolent response,” but it may also reflect more general prosocial behavior. For these reasons, rather than delete these three items, we chose to incorporate cross-loadings for these items into our model.

The addition of the three cross-loadings to the seven-factor model for Sample 1 (see Model 12 in Table 1) and to the eight-factor model for Sample 2 (see Model 14 in Table 2) resulted in models that fit the data well and did not result in any estimation problems. Within these models, the standardized factor loadings were all significant at $p < .001$ (see Supplemental Table S-1). As would be expected items allowed to load on two factors had lower loadings, but all were above .33 (i.e., .37 to .58). For items with cross loadings, all of the loadings on primary factors were higher than the cross-loadings for both waves of Sample 1. For Sample 2, two of these items had cross-loadings that were higher than loadings on the primary factors (i.e., .53 versus .43, and .46 versus .38). All remaining items had loadings ranging from .62 to .99. Only one or two of these loadings within each sample were less than .70, and over three-fourths were .80 or higher. Correlations among the factors within each wave are reported in Table 3 for Sample 1 and Table 4 for Sample 2. Although the correlations among the three aggression factors are lower than in the original models, they are still quite high. Table 3 also reports correlations across waves for Sample 1. These were highest for the three aggression factors (*rs* = .66 to .68), and slightly lower for the two victimization (*rs* = .54 to .56) and positive behavior factors (*rs* = .49 to .51).

Based on the high correlations among factors within the three domains we examined a model specifying higher-order factors representing overall aggression, victimization, and positive behavior. We based the model for Sample 1 on the seven-factor model that included cross-loadings (see Figure 1) assuming configural invariance such that the patterns of loadings for the seven first-order factors and for the second-order factors were the same across waves, but the estimates were allowed to differ. In addition to serial correlations for the items across waves, we included serial correlations across waves for residual variances for the first-order factors. The initial model for Sample 1 fit the data very well, but included a small nonsignificant negative estimate for the residual variance of one of the first-order factors. Constraining this to zero resulted in a higher-order factor model that fit the data as well as the less parsimonious first-order factor model based on comparison of the CFIs (see Model 13 in Table 1). We also examined a higher-order version of the model for Sample 2 that included eight first-order factors with cross-loadings. This model (see Model 15 in Table 2) also fit the data very well and resulted in minimal change in the fit indices relative to the first-order factor model. The findings for both samples suggest that little information was lost by representing relations among factors within each domain by higher-order factors. The higher-order model with cross-loadings was therefore the basis for all subsequent analyses.

Measurement Invariance Over Intervention Condition and Time in Sample 1

The data collected from Sample 1 provided an opportunity to evaluate measurement invariance across intervention conditions and over time. We conducted these analyses based on the seven-factor model that included cross-loadings and three higher-order factors. We used multiple group analyses of the Wave 2 data to test measurement invariance across two groups of participants – those in classrooms where the intervention has been implemented during the school year ($N = 256$), and those in the classrooms where it had not ($N = 404$). This model fit the data very well very well (see Model 15 in Table 1). We then tested a model that imposed strong measurement invariance by constraining item thresholds and loadings for the first-order and higher-order factors to the same values across groups. This model fit the data as well as the model specifying configural invariance based on the fit indices and results of the difference test (see Model 16 in Table 2). These results supported strong measurement invariance across groups that differed in their exposure to the intervention.

We conducted additional analyses based on the longitudinal higher-order factor model to test the stability of the factor structure over time. We compared the initial model that specified configural invariance (i.e., Model 13 in Table 1) to a model that specified strong measurement invariance (i.e., constrained the loadings and thresholds to the same values across waves) (see Models 14 in Table 1). Imposing strong measurement invariance had minimal impact on the fit of the model (i.e., ΛCFI < .001). Within this model, residual variances for first-order factors were significantly correlated over time ($rs = .36$ to .53 for the three aggression factors; .59 and .61 for the two victimization factors; and .38 for Prosocial Behavior). This suggests some stability in these factors not accounted for by the higher order factors. In contrast, the residual variance for the Effective Nonviolent Behavior factor was not significant at either wave or significantly correlated over time.

Measurement Invariance across Gender, Grade, and Intervention Condition

We ran multiple group analyses to investigate measurement invariance across gender, grade, and intervention status using data from Sample 2. We conducted these analyses on the eight-factor higher-order factor model that included cross-loadings. Comparison of multiple group models provided support for strong measurement invariance across gender. The model specifying configural invariance (Model 16) and the model specifying strong measurement invariance (see Model 17 in Table 2) both fit the data equally well. Imposing strong measurement invariance allowed us to compare means for boys and girls. Within this model boys had significantly higher scores than girls on the Aggression factor (Cohen's $d = .24, p < .001$) and Victimization factor (Cohen's $d = .30, p < .001$), and lower scores on the Positive Behavior factor (Cohen's $d = -.45, p < .001$).

Comparison of multiple group models also provided support for strong measurement invariance across grades. Both models fit the data very well. Within the strong measurement invariance model it was necessary to constrain a small negative variance for the residual variance in the Effective Nonviolent Behavior factor for the eighth grade to zero to obtain a proper estimate of model parameters. The strong measurement invariance model fit the data as well as the configural invariance model and resulted in a small improvement in the CFI. Within this model there were no significant differences in means for the Aggression, Victimization, or Positive Behavior factors across grades.

We also conducted tests of measurement invariance to determine if intervention activities influenced the measurement properties of the PBFS-TR. We used multiple group analyses to compare ratings of students in Sample 2 completed by teachers at a school during the time the intervention was being implemented to those completed in the absence of the intervention. The findings suggested that the presence of the intervention did not influence the structure or key parameters (i.e., loadings and item thresholds) of the measure (see models 20 and 21 in Table 2).

Relation between PBFS-TR Factors and SSIS Scores

We evaluated the convergent validity of scores on the PBFS-TR based on their pattern of relations with scales on the SSIS in Sample 1. We expanded the longitudinal model that included higher-order PBFS-TR factors under the assumption of strong measurement invariance to include scores on the SSIS scales as manifest variables. The resulting model fit the data very well (see Model 17 in Table 1). To simplify the interpretation of results, we also tested a model in which the relations (i.e., covariances) between SSIS scales and PBFS factors were constrained to the same values across waves. Imposing this constraint did not result in a significant decrease in model fit based on the difference test and comparison of fit indices (see Model 18 in Table 1). This indicates that relations between PBFS-TR factors and SSIS scores within each wave did not differ over time.

Correlations among the three PBFS-TR higher-order factors and the six SSIS scales are reported in Table 5. As hypothesized, the PBFS-TR Aggression factor was highly correlated with the SSIS Bullying scale ($r = .68$) and negatively correlated with SSIS Self-Control, Responsibility, and Empathy scales ($rs = -.64$ to $-.48$, respectively). It also had a moderate positive correlation with SSIS Internalizing Problems ($r = .27$) and negative correlation with Academic Competence ($r = -.30$). Support was also found for hypotheses regarding relations between the PBFS-TR Positive Behavior factor and SSIS scales. The PBFS-TR Positive Behavior factor had large positive correlations with the three SSIS social skills scales (i.e., Self-Control, Responsibility, Empathy) ($rs = .69$ to $.75$) and with SSIS Academic Competence scale ($r = .52$), and was negatively correlated with the Bullying scale ($r = -.51$). The PBFS-TR Victimization factor displayed a similar pattern of correlations with the SSIS Bullying and the three social skills scales, but with somewhat lower values than those found for the Aggression factor. Consistent with hypotheses, of the three PBFS-TR higher-order factors the Victimization factor had the strongest correlation with the SSIS Internalizing Problems scale ($r = .41$).

We also examined patterns of change across waves for PBFS-TR and SSIS scales. The PBFS-TR Aggression factor was highly correlated across waves ($r = .72$), as was the SSIS Bullying scale ($r = .61$). We calculated effect sizes by dividing the differences in means across waves by the standard error of the difference score (Gibbons, Hedeker, & Davis, 1993). Teachers ratings on both the PBFS-TR Aggression and SSIS Bullying scale increased from the beginning to the end of the school year with a higher increase on the PBFS-TR ($d_{rm} = .54$, $p < .001$) than on the SSIS ($d_{rm} = 25$, $p = .022$). The PBFS-TR Victimization factor was also highly correlated across waves ($r = .53$), with scores increasing across the school year ($d_{rm} = .56$, $p < .001$). The correlation between scores on the PBFS-TR Positive

Behavior factor across waves *(r* = .58) was within the range of correlations across waves found for the SSIS social skills scales (*rs* = .54 to .63). However, whereas scores on two of the three SSIS social skills scales showed a slight *decrease* across waves (d_{rm} = −.11, *p* = .041 for Self-Control, and d_{rm} = −.16, *p* = .016 for Responsibility), scores on the PBFS-TR Positive Behavior factor showed a slight *increase* (d_{rm} = .23, *p* = .011).

Relation between PBFS Teacher and Student Reports

We also examined the convergent validity of scores on the PBFS-TR based on its relations with student self-ratings on the PBFS-AR in Sample 2. We incorporated the PBFS-AR and Intentions for Nonviolent Behavior scale into the model that included the higher-order PBFS-TR factors. Within these models, we used item-level data to create latent variables representing the factors on the PBFS-AR and Intentions for Nonviolent Behavior scale treating all items as ordered categorical variables. The resulting model fit the data very well (see Model 22 in Table 2). Correlations among the resulting factors are reported in Table 6. As hypothesized, the PBFS-TR Aggression factor was positively correlated with the PBFS-AR Aggression, Delinquent Behavior, and Substance Use factors (*rs* = .17 to .24), and was inversely related to the Intentions for Nonviolent Behavior factor (*r* = −.38). The PBFS-TR Victimization factor was positively correlated with the PBFS-AR Victimization (*r* = .18) factor, and with the Aggression, Delinquent Behavior, and Substance Use factors (*rs* = .09 to .18), and was negatively correlated with the Intentions for Nonviolent Behavior factor (*r* = −.24). The PBFS-TR Positive Behavior factor also demonstrated the expected pattern of correlations with the Intentions for Nonviolent Behavior factor (*r* = .29), and the PBFS-AR problem behavior factors (*rs* = −.19 to −.24).

Although all correlations were significant and in the expected directions, evidence for discriminant validity was at best mixed. For example, the PBFS-TR Aggression factor was more strongly related to the PBFS-AR Aggression factor than to the PBFS-AR Victimization factor (*p* < .001), but had a stronger (in absolute value) relation to student reports on the Effective Nonviolent Behavior scale (*p* = .002). Looking across row S04 in Table 6, student ratings of their frequency of victimization on the PBFS-AR had the highest correlation with teacher ratings of victimization on the PBFS-TR than with other teacher ratings (*p* < .01). However, looking down column S-4 this correlation did not differ in magnitude from correlations between PBFS-TR Victimization factor and student ratings of their aggression, delinquent behavior, or nonviolent positive behaviors. Finally, the PBFS-TR Positive Behavior factor had its strongest cross-method correlation with student ratings of their intentions to use effective nonviolent behaviors. This correlation was significantly stronger than cross-method correlations with student ratings of their frequency of aggression (*p* = .04) and victimization (*p* < .001), but it had equally strong correlations with delinquent behavior and substance use.

We conducted sensitivity analyses to determine if the results were influenced by missing data on the Intentions for Nonviolent Behavior scale, which was not administered during Year 5. Reanalysis of the data restricting the sample to participants with data from the first four years of the project resulted in fit statistics that were within .001 of those obtained with the larger sample. With the exception of one correlation that differed by .04, all other correlations between the Intentions for Nonviolent Behavior scale and other measures based on the reanalysis were within .02 of those found with the larger sample. We choose to report the analyses based on the full sample in order to make full use of the available data on all other measures.

Discussion

Overall, the results of this study supported the PBFS-TR as a teacher-report measure of adolescents' aggression, victimization, and positive behaviors. Its hypothesized factor structure fit the data well and fit significantly better than several competing models. We found evidence of strong measurement invariance across time, middle school grades, gender, and for students that differed in their exposure to violence prevention programs. The pattern of factor means was generally consistent with gender differences in aggression and victimization reported in other studies using adolescent self-report (e.g., Farrell et al., 2016) and studies examining gender differences in prosocial behavior (Fabes, Carlo, Kupanoff, & Laible, 1999). There was also support for the concurrent validity of the PBFS-TR based on its correlations with teacher ratings on SSIS scales and student ratings on the PBFS-AR and a measure of behavioral intentions.

Our analysis of the PBFS-TR found support for separate factors representing physical, verbal, and relational forms of aggression. The ability of teachers to differentiate among these forms of aggression has not previously been demonstrated, despite evidence supporting these distinctions within adolescent reported data (Card et al., 2008). It is, however, important to note that there were strong correlations among specific forms of aggression. There was also support for a higher-order factor that subsumed all three forms of aggression. We found the weakest support was for differentiating verbal aggression from physical and relational aggression. Models in which verbal aggression was combined with either physical aggression or relational aggression fit the data nearly as well as the model specifying three separate factors. In contrast, there was stronger support for differentiating between physical and relational aggression based on the lower correlation between these two measures, and the decrease in model fit when they were combined into a single factor along with verbal aggression. This provides some support for the notion that teachers can differentiate between readily observable behaviors such as physical aggression and more subtle forms such as relational aggression.

There are several factors that may account for these high correlations among teacher ratings of different forms of aggression. One possibility is that they reflect attributional biases wherein teachers may tend to be influenced not only by the behavior they observe, but also by their perceptions of students' dispositions (De Los Reyes & Kazdin, 2005). Anchoring the PBFS-TR to a timeline (i.e., the past 30 days) and focusing on behavioral frequency may have helped reduce teachers' tendency to use a global approach. Nonetheless, although we believe there is value in focusing teachers' ratings on examples of specific behaviors, we cannot rule out the possibility that teachers were influenced not only by the behaviors they observed, but also by their overall perceptions of the students they rated. Another possibility is that the behaviors the items are designed to assess are highly likely to co-occur. Adolescents who engage in physical forms of aggression may also be very likely to engage in verbal acts such as name calling, teasing, and insulting other adolescents. In general, the pattern of high correlations among the different forms of aggression on the teacher form of the PBFS is consistent with a study by Farrell et al. (2016) who found a similar pattern in their analysis of the structure of the adolescent report form of the PBFS.

There was clearer evidence to support separate factors representing specific forms of victimization. Analyses of Sample 1 indicated that a two-factor model that differentiated between physical and relational forms of victimization fit the data significantly better than a model that specified a single Victimization factor. Analyses of an expanded version of the PBFS-TR administered to Sample 2 found clear support for a three-factor model that differentiated among physical, verbal, and relational forms of victimization. Nonetheless, there were high correlations among all three forms of victimization, though they were not as extreme as some of those found among factors representing specific forms of aggression. There was also support for a higher-order Victimization factor.

Whether there is a benefit in differentiating among forms of aggression and victimization (i.e., physical, verbal, relational) may depend on the purpose. Prior research has revealed differences in the prevalence rates, causes, and the consequences of different forms of aggression (see review by Card et al., 2008). This suggests that the use of global measures of aggression and victimization may fail to detect factors that are not common to all of these forms. Measures of specific forms may also be useful for evaluating prevention efforts. For example, a recent study showed a sequential impact on different forms of both aggression and victimization after implementation of a school-based violence prevention program, such that effects on verbal and relational aggression and victimization emerged earlier than effects on physical aggression and victimization (Author reference). The use of global or higher-order measures would have masked or possibly missed earlier impacts of the intervention.

This study supported the notion that prosocial behavior and effective nonviolent behavior, although related, are not identical constructs. Teachers differentiated between prosocial behaviors (e.g., tried to do their best in school, helped out around the school or classroom) and effective non-violent strategies (e.g., solved a disagreement peacefully, walked away when someone wanted to fight). This suggests there may be value in identifying the factors that influence adolescents’ use of non-violent strategies that effectively solve interpersonal conflicts. Further study is needed to examine whether the cultivation of prosocial behavior and effective nonviolent behavior predict different adjustment patterns.

In general, this study provided support for the convergent validity of scores on the PBFS-TR based on its pattern of correlations with a well-established teacher report measure, the SSIS (Gresham & Elliott, 2008). As hypothesized, teacher ratings of both aggression and victimization were highly correlated with bullying as measured by the SSIS, as is consistent with other research during early adolescence (e.g., Haynie et al., 2001). Furthermore, the correlations between the PBFS-TR and SSIS scales measuring empathy, self-control, and responsibility were consistent with other empirical work. In a review of 17 studies, Lovett and Sheffield (2007) identified a robust relation between empathy and aggression, specifically within self-report measures of both. Our results further support this connection within teacher ratings of behavior. Others have also found that self-reported measures of social skills have been negatively related to measures of aggression (e.g., Bussey, Quinn, & Dobson, 2015) and positively related to measures of prosocial behavior (see meta-analysis by Eisenberg & Miller, 1987) and effective nonviolent behavior. Teacher ratings on both the PBFS-TR and SSIS suggested increases in students' aggressive behavior across the school year, with greater increases found on the PBFS-TR compared with the SSIS. The Aggression subscale of the PBFS-TR assessed multiple forms of aggression compared with the narrower focus of bullying on the SSIS. The focus on specific behaviors and broader range of coverage on the PBFS-TR may have made it more sensitive to change. This may reflect increases in students' frequency of aggressive behavior during the school year. It is also possible that teachers become better at identifying instances of aggression as they become more acquainted with their students.

We also found support for the validity of the PBFS-TR for assessing adolescents' behavior based on its relations with student ratings on the PBFS-AR. Specifically, teacher and student ratings of related constructs were significantly correlated, suggesting consistency across informants. Although correlations between adolescent and teacher ratings were low to moderate, they were within the range typically found for measures from different informants (see review by Meyer et al., 2001). This low level of agreement is not solely a function of measurement error and informant bias. It also reflects important differences in the context in which behavior occurs and is observed. Teachers' observations of students are limited to the school context and to situations in which they are present. School sanctions for aggression and disruptive behavior make it less likely to occur in school, especially in situations where teachers are present to observe it. The fact that each source of data is subject to different types of bias reinforces the value of obtaining reports from multiple sources (De Los Reyes, & Kazdin, 2005). One surprising finding was that teacher ratings were most strongly correlated with students' ratings of their behavioral intentions. One possible explanation for this finding is that students' reports of their intentions or likelihood of reacting to a peer may reflect their own assessment of their disposition more so than their recollection of how they usually react in those situations. In that sense, students' reports of their intentions may be more likely to reflect the tendency of observers to overemphasize dispositions and underestimate situational factors (De Los Reyes & Kazdin, 2005).

Our analyses of data from two projects in which participants varied in their degree of exposure to violence prevention programs provided an opportunity to determine whether the interventions influenced the structure of teachers' ratings on the PBFS-TR. In Sample 1, we evaluated measurement invariance across groups of students in the same school who differed in whether they were in classrooms that implemented the Second Step curriculum (Committee for Children, 2008). Our analyses of Sample 2 tested measurement invariance across groups of students in schools that differed in whether the Olweus Bullying Prevention Program (Olweus & Limber, 2010) was being implemented. These are distinct approaches to violence prevention. Whereas Second Step involves a classroom level curriculum, the Olweus Bully Prevention Program is a school-level intervention that includes individual-level, classroom-level, and school-level components. Because both approaches involve teachers it is plausible that their involvement could influence how they perceive behaviors related to aggression and victimization. For example, teachers may become more adept at differentiating among forms of aggression or more sensitive to subtle forms of victimization (e.g, relational aggression). The current study found support for strong measurement invariance across intervention conditions, indicating that although teachers were actively involved in these interventions, it did not affect the structure of their ratings on the PBFS-TR. This is an important finding in that intervention effects on the measure would have compromised analyses comparing scores across conditions or pre-to-post comparisons. Measurement invariance across intervention conditions is an important, but seldom studied property of measures given the frequency with which measures are used in schools where interventions are being implemented. Further work is needed to determine whether similar effects might be observed on other measures of student behavior.

Limitations and Future Research

This study had several limitations that warrant discussion. Because the participants were from a primarily African American sample of students from an urban school system, the findings may not generalize to other populations of adolescents. It is also unclear whether the current findings would generalize to other age ranges, in that older adolescents may make more of an effort to hide problem behaviors from their teachers (Achenbach et al., 1987). We evaluated the convergent validity of scores on the PBFS-TR based on their pattern of correlations with teacher ratings on the SSIS and with student reports on the PBFS-AR. Both are imperfect criteria. Although the SSIS is widely used, it is subject to the same limitations and sources of bias as other teacher rating scales. There was also not a perfect match between the content of scales on the PBFS-TR and SSIS, making it difficult to evaluate convergent and discriminant validity. In contrast, the PBFS-AR was based on student reports of their behavior, which are subject to the biases commonly found with self-report measures. As previously noted, its focus was not limited to behavior at school, which may have attenuated its relation to a teacher report measure. Further work is needed to evaluate the validity and utility of using the PBFS-TR to assess adolescents' behavior in other samples and to determine its relation to other criteria.

Overall, our findings support the notion that teachers can differentiate among various forms of aggression, victimization, and prosocial behaviors, and that the PBFS-TR can provide useful data to assess students' experiences within the school setting. The current analyses suggest that the PBFS-TR can also provide broad measures of aggression, victimization, and positive behaviors. These may be helpful for school administrators and clinicians who are concerned with the general level of each construct rather than specific forms. In either case, we recommend researchers using the PBFS-AR exclude the four items that contributed minimal information to reliability or that did not fit in with the overall structure, and include the additional items representing verbal victimization that were administered to Sample 2 (see Supplemental Table S1). In addition to providing a first-order factor representing verbal victimization, it expands the domain captured by the higher-order Victimization factor. We also recommend scoring the measure by constructing latent variables for each construct based on models like those in the current study that treat items as ordered categorical variables. That approach takes into account differences in the severity of individual items (e.g., does not treat being in a physical fight and pushing or shoving someone as equally serious), and does not assume items are measured on an equal interval scale. It also allows for including cross-loadings for the three items we identified based on both substantive considerations and item analysis. Moreover, this strategy has the other advantages of latent variable models such as providing an explicit test of the measurement model and producing unbiased estimates of relations among latent constructs by explicitly modeling measurement error (Bollen, 1989).

Although using items on the PBFS-TR to create latent variables is preferred, we acknowledge that there may be circumstances that require that researchers use a less complex approach to scoring. In particular, treating the items as ordered categorical variables requires a sufficiently large sample to ensure that there are not cells with zeros or very small numbers of observations. Even with the large sample sizes in the current study, we found it necessary to merge some categories that had very low frequencies. A simpler, though less desirable scoring approach is to create manifest variables by averaging items within each scale. In this case combining categories with low frequencies is not necessary because they do not create estimation issues. In calculating such scores, it is advisable to include items only on the scale where it has its primary loading. Although including cross-loadings in a factor model can reduce the correlations between factors (Marsh et al., 2009), it has the opposite effect when scores are directly calculated. One cautionary note is that our finding that two of the items had slightly larger cross-loadings than their primary loadings in one of the samples suggests the need for further work to verify where these items may best be represented in the overall structure. We evaluated the impact of a more direct approach to scoring by correlating scores based on averaging items within each scale with estimated factor scores on the corresponding scale using data from Sample 2. The resulting correlations ranged from .94 to .99. This suggests that directly calculating scores may be a feasible alternative when it is not possible to use the preferred method of constructing latent variables from the item-level data.

Prior research has shown the importance of distinguishing between forms of aggression in terms of their prevalence, causes, and consequences (e.g., Card et al., 2008), and teachers can provide a cost-effective way in obtaining this information (Clemans et al., 2014). Problem and prosocial behaviors do not appear to be extremes on the same continuum, but rather unique constructs that allow youth to engage in both risky and prosocial behaviors (Orpinas et al., 2015). These findings also have important implications for school-based violence prevention efforts. As researchers increasingly advocate for concurrent efforts to reduce problem behaviors and support nonviolent and prosocial behaviors (e.g., Farrell et al., 2007; Greenberg et al., 2003), the PBFS-TR provides a tool for assessing multiple domains simultaneously within the school setting. Teachers are in a unique position to evaluate school-wide intervention efforts that incorporate bullying prevention (e.g., Olweus & Limber, 2010) with positive behavior interventions, such as empathy training (e.g., Sahin, 2012) and social-emotional learning modules (e.g., Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, CDC Cooperative Agreement 5U01CE001956 and a grant from the Institute of Education Sciences (R324A100160). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the Institute of Education Sciences.

References

Achenbach TM, McConaughy SH, & Howell CT (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101(2), 213–232. doi:10.1037/0033-2909.101.2.213 [PubMed: 3562706]

Asparouhov T & Muthen B (2009). Exploratory structural equation modeling. Structural. Equation Modeling, 16:397–438. doi:/ 10.1080/10705510903008204

Bollen KA (1989). Structural equation modeling with latent variables. New York: Wiley.

Bussey K, Quinn C, & Dobson J (2015). The moderating role of empathic concern and perspective taking on the relationship between moral disengagement and aggression. Merrill-Palmer Quarterly, 61, 10–29. doi:10.13110/merrpalmquar1982.61.1.0010

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(5), 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(2), 233–255. doi:10.1207/S15328007SEM0902_5

Clemans KH, Musci RJ, Leoutsakos JS, & Ialongo NS (2014). Teacher, parent, and peer reports of early aggression as screening measures for long-term maladaptive outcomes: Who provides the most useful information? Journal of Consulting and Clinical Psychology, 82(2), 236–247. doi:10.1037/a0035918 [PubMed: 24512126]

Committee for Children. (2008). Second Step: Student success through prevention program. Seattle, WA: Committee for Children.

Crick NR (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. Child Development, 67(5), 2317–2327. doi:10.2307/1131625 [PubMed: 9022243]

Cullerton-Sen C, & Crick NR (2005). Understanding the effects of physical and relational victimization: The utility of multiple perspectives in predicting social-emotional adjustment. School Psychology Review, 34(2), 147–160.

Dahlberg L, Toal S,. Swahn M, & Behrens C. (2005) Measuring Violence-Related Attitudes, Behaviors, and Influences Among Youths: A Compendium of Assessment Tools, 2nd ed Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

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]

Durlak JA, Weissberg RP, Dymnicki AB, Taylor RD, & Schellinger KB (2011). The impact of enhancing students' social-emotional learning: A meta-analysis of school-based universal interventions. Child Development, 82, 405–432. doi:10.1111/j.1467-8624.2010.01564.x [PubMed: 21291449]

Eisenberg N, & Miller PA (1987). The relation of empathy to prosocial and related behaviors. Psychological Bulletin, 101, 91–119. doi:10.1037/0033-2909.101.1.91 [PubMed: 3562705]

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

Fabes RA, Carlo G, Kupanoff K, & Laible D (1999). Early adolescence and prosocial/moral behavior I: The role of individual processes. Journal of Early Adolescence, 19, 5–16. doi:10.1177/0272431699019001001

Farrell AD, Erwin EH, Allison KW, Meyer A, Sullivan T, Camou S, . . . Esposito L (2007). Problematic situations in the lives of urban African American middle school students: A qualitative study. Journal of Research on Adolescence, 17(2), 413–454. doi:10.1111/j.1532-7795.2007.00528.x

Farrell AD, Sullivan TN, Goncy EA, & Le AH (2016). Assessment of adolescents' victimization, aggression, and problem behaviors: Evaluation of the Problem Behavior Frequency Scale. Psychological Assessment, 28(6), 702–714. doi:10.1037/pas0000225 [PubMed: 26372261]

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, NY: Routledge.

Gibbons RD, Hedeker DR, & Davis JM (1993). Estimation of effect size from a series of experiments involving paired comparisons. Journal of Educational Statistics, 18, 271–279. doi.org/10.2307/1165136

Goldfried MR & D'Zurilla TJ (1969). The behavioral-analytic method for assessing competence In Spielberger CD (Ed.), Current topics in clinical and community psychology (pp. 151–196). New York: Academic Press.

Greenberg MT, Weissberg RP, O'Brien MU, Zins JE, Fredericks L, Resnik H, et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. American Psychologist, 58, 466–474. doi: http://dx.doi.org.proxy.library.vcu.edu/10.1037/0003-066X.58.6-7.466 [PubMed: 12971193]

Gresham FM, Cook CR, Crews SD, & Kern L (2004). Social skills training for children and youth with emotional and behavioral disorders: Validity considerations and future directions. Behavioral Disorders, 30(1), 32–46.

Gresham FM, & Elliott SN (2008). Social Skills Improvement System: Rating scales manual. Minneapolis, MN: Pearson Assessments Retrieved from http://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssis-rating-scales.html

Haynie DL, Nansel T, Eitel P, Crump AD, Saylor K, Yu K, & Simons-Morton B (2001). Bullies, victims, and bully/victims: Distinct groups of at-risk youth. Journal of Early Adolescence, 21, 29–49. doi: 10.1177/0272431601021001002

Hu L, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55. doi: 10.1080/10705519909540118

Lovett BJ, & Sheffield RA (2007). Affective empathy deficits in aggressive children and adolescents: A critical review. Clinical Psychology Review, 27, 1–13. doi:10.1016/j.cpr.2006.03.003 [PubMed: 16697094]

Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJS, & Trautwein U (2009) Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. Structural Equation Modeling: A Multidisciplinary Journal, 16, 439–476, DOI: 10.1080/10705510903008220

Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, Eisman EJ, Kubiszyn TW, & Reed GM (2001). Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist, 56, 128–165. doi: http://dx.doi.org.proxy.library.vcu.edU/10.1037/0003-066X.56.2.128 [PubMed: 11279806]

Muthén B, & Asparouhov T (2006). Robust chi square difference testing with mean and variance adjusted test statistics. Mplus Web Note No. 10. Retrieved from: https://www.statmodel.com/download/webnotes/webnote10.pdfonFebruary 24, 2017.

Muthén LK & Muthén BO (2015). Mplus User's Guide Seventh Edition Los Angeles, CA: Author.

Muthén BO & Satorra A (1995). Complex sample data in structural equation modeling. Sociological Methodology, 25: 267–316. doi:/ 10.2307/271070

Olweus D, & Limber SP (2010). Bullying in school: Evaluation and dissemination of the Olweus bullying prevention program. American Journal of Orthopsychiatry, 50(1), 124–134. doi:10.1111/j.1939-0025.2010.01015.x

Orpinas P, Raczynski K, Peters JW, Colman L, & Bandalos D (2015). Latent profile analysis of sixth graders based on teacher ratings: Association with school dropout. School Psychology Quarterly, 30(4), 577–592. doi:10.1037/spq0000107 [PubMed: 25528590]

Reijntjes A, Kamphuis JH, Prinzie P & Telch MJ (2010). Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. Child Abuse and Neglect, 34, 244–252. doi: 10.1016/j.chiabu.2009.07.009 [PubMed: 20304490]

Reynolds CR, & Kamphaus RW (2004). Behavioral Assessment System for Children-Second Edition. Minneapolis, MN: Pearson Assessments. Retrieved from http://www.pearsonclinical.com/education/products/100000658/behavior-assessment-system-for-children-second-edition-basc-2.html

Sahin M (2012). An investigation into the efficiency of empathy training program on preventing bullying in primary schools. Children and Youth Services Review, 34, 1325–1330. doi: 10.1016/j.childyouth.2012.03.013

Vitaro F, Brendgen M, Girard A, Dionne G, Tremblay RE, & Boivin M (2016). Links between friends' physical aggression and adolescents' physical aggression: What happens if gene-environment correlations are controlled? International Journal of Behavioral Development, 40(3), 234–242. doi: 10.1177/0165025415618277

Wentzel KR, Filisetti L, & Looney L (2007). Adolescent prosocial behavior: The role of self-processes and contextual cues. Child Development, 78(3), 895–910. doi: 10.1111/j.1467-8624.2007.01039.x [PubMed: 17517011]

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 ofprevention: Methodological advances from alcohol and substance abuse research (pp. 281–324). Washington, DC, US: American Psychological Association. doi :10.1037/10222-009

Author Manuscript

Author Manuscript

Author Manuscript

Public Significance Statement:

This study found evidence supporting the use of a measure teachers can use to rate aggressive behaviors, prosocial and nonviolent behaviors, and victimization experiences of middle school students.

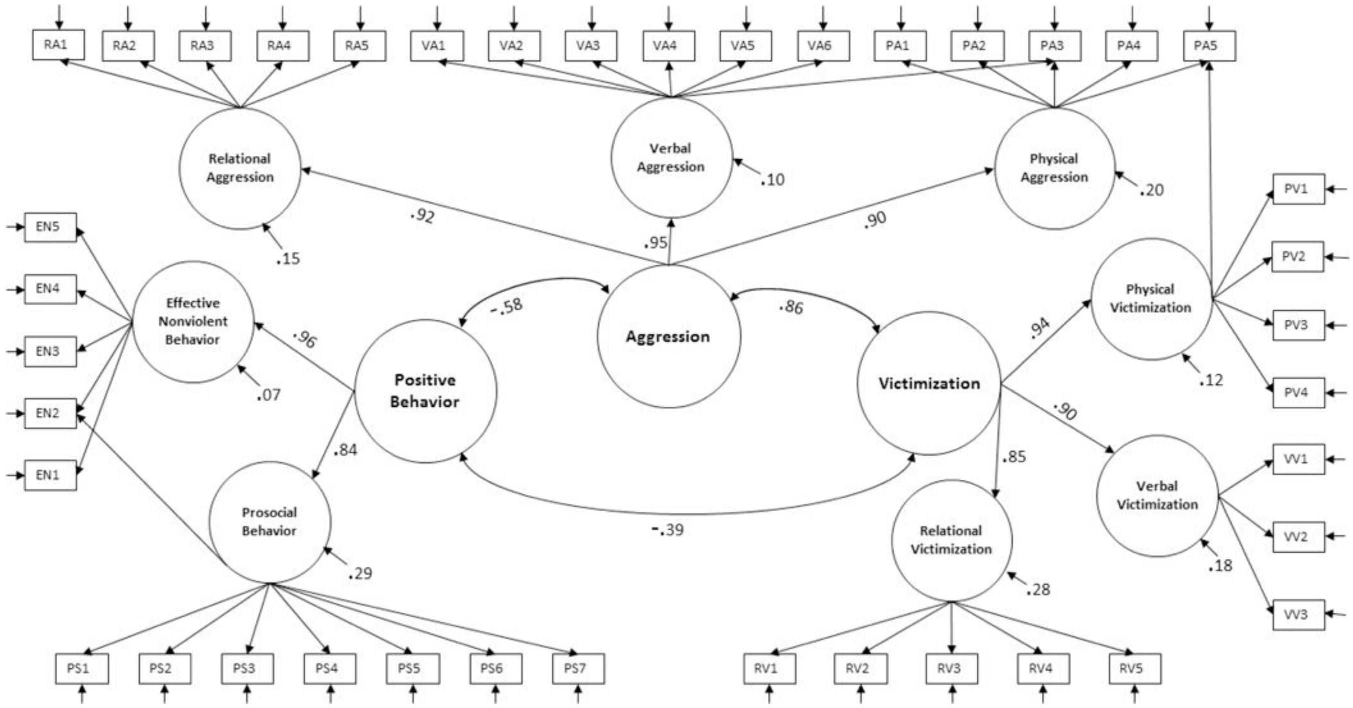

Figure 1:

Structural model of the Problem Behavior Frequency Scale – Teacher Report specifying eight first-order factors and three higher-order factors. Parameter estimates are standardized loadings for higher-order factors, correlations, and residuals for first-order factors from Sample 2. The figure does not display the factor loadings and measurement errors for the first-order factors.

Table 1

Fit indices for models of the structure of the Problem Behavior Frequency Scale-Teacher Report scale in Sample 1

Model	Description	Model χ^2	df	RMSEA	CFI	TLI	Comp	ΔCFI	$\chi^2\Delta^a$	df
	Aggression submodels									
1	3 factors: Physical, Verbal, & Relational	562.96***	375	.026	.993	.992				
2	2 factors: Overt & Relational	717.44***	384	.035	.987	.986	1	−.006	89.44***	9
3	2 factors: Physical & Nonphysical	753.57***	384	.036	.986	.984	1	−.007	124.69***	9
4	1 factor: Overall Aggression	950.50***	389	.045	.979	.979	1	−.014	201.08***	14
	Victimization submodels									
5	2 factors: Physical & Relational	363.11***	120	.053	.980	.975				
6	1 factor: Overall Victimization	559.7***	125	.069	.965	.957	5	−0.15	108.31***	5
	Positive behavior submodels									
7	2 factors: Prosocial & Effective Nonviolent	879.33***	330	.048	.941	.933				
8	1 factor: Positive Behavior	1069.7***	335	.055	.921	.911	7	−.020	160.29***	5
9	2 factors: 2 items deleted	439.16***	234	.035	.976	.972				
10	1 factor: 2 items deleted	651.00***	239	.049	.952	.944	9	−.024	128.39***	5
	Full 7-factor models									
11	Without cross-loadings	3090.30***	2357	.021	.978	.977	12	−.002	77.70***	6
12	With cross-loadings	3029.48***	2351	.020	.980	.978				
13	With higher-order factors	3097.31***	2408	.020	.980	.979	12	.000	127.16***	57
14	With higher-order factors & strong invariance	3191.41***	2495	.020	.980	.979	13	.000	123.73**	87
	Multiple-group analysis of Wave 2 7-factor higher-order model by intervention condition (N = 660)									
15	Configural invariance	1703.73***	1163	.038	.982	.980	16	.000	88.50	88
16	Strong measurement invariance	1769.10***	1251	.035	.982	.982				
	Validity model with Social Skills Improvement System									

Model	Description	Model χ^2	df	RMSEA	CFI	TLI	Comp	ΔCFI	$\chi^2\Delta^a$	df
17	Initial validity model	4121.63***	3287	.019	.975	.975				
18	Model with constrained	4113.09***	3305	018	.976	.974	17	.001	22.03	18

Note. All models N = 727 and longitudinal except where noted. RMSEA = root mean-square error of approximation. CFI = comparative fit index. TLI = Tucker-Lewis fit index.

[a]Difference test. Significant chi-square values indicate a significant decrease in fit relative to the comparison model (Comp).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 2

Fit indices for competing models of the factor structure of the Problem Behavior Frequency Scale-Teacher Report in Sample 2

Model	Description	Model χ^2	df	RMSEA	CFI	TLI	Comp	Δ CFI	$\chi^2\Delta$	df
	Aggression submodels									
1	3 factors: Physical, Verbal, & Relational	478.30***	101	.046	.991	.989				
2	2 factors: Overt & Relational	765.82***	103	.061	.984	.981	1	−.007	147.76***	2
3	2 factors: Physical & Nonphysical	812.95***	103	.063	.983	.980	1	−.008	129.62***	2
4	1 factor: Overall Aggression	1166.32***	104	.077	.974	.970	1	−.017	271.04***	3
	Victimization submodels									
5	3 factors: Physical, Verbal, & Relational	483.80***	51	.070	.974	.966				
6	2 factors: Overt & Relational	689.06***	53	.083	.962	.952	5	−.012	120.24***	2
7	2 factors: Physical & Nonphysical	777 14***	53	.089	.956	.946	5	−.018	172.36***	2
8	1 factor: Overall Victimization	1045.55***	54	.103	.940	.927	5	−.034	288.49***	3
	Positive behavior submodels									
9	2 factor: Prosocial & Effective Nonviolent	1310.97***	76	.097	.903	.884				
10	1 factor: Positive Behavior	1938.04***	77	.118	.854	.828	9	−.049	253.8***	1
11	2 factor: 2 items deleted	479 12***	53	.068	.964	.955				
12	1 factor: 2 items deleted	835.39***	54	.091	.934	.920	11	−.030	114.16***	1
	Full Eight-factor models									
13	8 factors without cross-loadings	2422.30***	712	.037	.962	.959	14	.003	146.67***	3
14	8 factors with cross-loadings	2289.97***	709	.036	.965	.962				
15	8 factors with higher-order factors	2544.48***	726	.038	.960	.957	14	−.005	256.46***	17
	Multiple group analysis of 8-factor models with higher-order factors									
16	Configurai invariance across gender	3284.95***	1452	.038	.969	.967			338.98***	34
17	Strong measurement invariance across gender	3296.87***	1550	.036	.970	.970	16	.000	97.28	98
18	Configurai invariance across grades	3871.51***	2178	.037	.968	.966			372 97***	51

Model	Description	Model χ^2	df	RMSEA	CFI	TLI	Comp	Δ CFI	$\chi^2\Delta$	df
19	Strong measurement invariance across grades	3938.47***	2375	.034	.971	.971	18	.003	187.53	197
20	Configurai invariance across conditions	3284.95***	1452	.038	.969	.967			338.98***	34
21	Strong measurement invariance across conditions	3296.87***	1550	.036	.970	.970	20	.001	97.28	98
		Validity with student ratings								
22	Validity model	4469.81***	3198	.015	.978	.977				

Note. N = 1,740. RMSEA = root mean-square error of approximation. CFI = comparative fit index. TLI = Tucker-Lewis fit index.

[a] Difference test. Significant chi-square values indicate a significant decrease in fit relative to the comparison model (Comp).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 3

Correlations among factors in the seven-factor model for Sample 1 at Wave 1 (above diagonal) and Wave 2 (below diagonal) and correlations across waves (on diagonal)

Factor	1	2	3	4	5	6	7
1. Physical Aggression	**.68*****	.85***	.77***	.77***	.61***	−.36***	−47***
2. Verbal Aggression	.88***	**.67*****	.87***	.65***	.63***	−43***	− 5***
3. Relational Aggression	.79***	.86***	**.66*****	.66***	.78***	−.32***	−.39***
4. Physical Victimization	.83***	.69***	.67***	**.54*****	.75***	−.21**	−28***
5. Relational Victimization	.59***	.61***	.76***	.69***	**.56*****	−.14	−.18*
6. Prosocial Behavior	−.50***	−53***	−.49***	−.38***	−.36***	**.49*****	.65***
7. Effective Nonviolent Behavior	−.63***	−.62***	−.58***	−51***	−.39***	.80***	**.51*****

Note. N = 727.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 4

Correlations among factors on the Problem Behavior Frequency Scale – Teacher Report for the eight-factor model for Sample 2

Factor	1	2	3	4	5	6	7
1. Physical Aggression							
2. Verbal Aggression	.86						
3. Relational Aggression	.78	.90					
4. Physical Victimization	.86	.75	.70				
5. Verbal Victimization	.65	.75	.68	.84			
6. Relational Victimization	.59	.65	.81	.74	.85		
7. Prosocial Behavior	−.44	−.50	−.37	−.35	−.31	−.20	
8. Effective Nonviolent Behavior	−.51	−.57	−.42	−.42	−.34	−.23	.81

Note. $N = 1{,}740$. All correlations are significant at $p < .001$.

Table 5

Means and correlations between Problem Behavior Frequency Scale – Teacher Report (PBFS-TR) higher-order factors and Social Skills Improvement System (SSIS) scales for Sample 1

	1	2	3	4	5	6	7	8	9
Problem Behavior Frequency Scale – Teacher Report									
1. Aggression	.72***								
2. Victimization	.87	.53***							
3. Positive Behavior	−.56***	−.28***	.58***						
Social Skills Improvement System									
4. Bullying	.68***	.54***	−.51***	.61***					
5. Self-Control	−.64***	−.47***	.69***	−.56***	.63***				
6. Responsibility	−.62***	−.45***	.75***	−.58***	.85***	.57***			
7. Empathy	−.48***	−.33***	.75***	−.46***	.74***	.80***	.54***		
8. Academic Competence	−.30***	−.23***	.52***	−.30***	.43***	.60***	.56***	.55***	
9. Internalizing Problems	.27***	.41***	−.30***	.39***	−.29***	−.35***	−.32***	−.34***	.49***
Wave 1 Mean	0.00	0.00	0.00	6.61	18.22	17.09	15.24	21.42	10.32
Wave 2 Mean	0.39	0.52	0.22	7.28	17.68	16.45	15.43	21.17	11.24
d_{rm}-coefficient[a]	0.54***	0.56***	0.23*	0.25*	−0.11*	−0.16*	0.05	−0.04	0.26***

N = 727. Note. Correlations are from Wave 1 based on a model in which the covariances between the PBFS-TR and SSIS within each wave were constrained to the same values across waves.

[a] d-coefficients are based on the difference between Wave 1 and Wave 2 means divided by the standard error of the difference scores.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 6

Correlations between higher-order factor scores on the teacher and student report forms of the Problem Behavior Frequency Scale and Intentions for Nonviolent Behavior scale for Sample 2.

	T-1	T-2	T-3	S-1	S-2	S-3	S-4
Teacher report							
T-1 Aggression							
T-2 Victimization	.86***						
T-3 Positive Behavior Student Report	−.58***	−.38***					
S-1 Aggression	.24***	.18***	−.19***				
S-2 Delinquent Behavior	.22***	.17***	−.24***	.86***			
S-3 Substance Use	.17***	.09*	−.19***	.63***	.73***		
S-4 Victimization	.08*	.18***	−.05	.65***	.51***	.32***	
S-5 Intentions for Nonviolent Behavior	−.38***	−.24***	.29***	−.19***	−.22***	−.19***	.11***

Note. $N = 1{,}740$.

* $p < .05$.

** $p < .01$.

*** $p < .001$.