J Psychopathol Behav Assess
DOI 10.1007/s10862-017-9623-9

Testing a Bifactor Model of Relational and Physical Aggression in Early Childhood

Kristin J. Perry[1] · Jamie M. Ostrov[1]

Abstract Different approaches have been used to classify children as relationally aggressive, physically aggressive, or both relationally and physically aggressive (co-occurring aggression). The goal of the current study ($N = 164$, 50.9% boys, $M_{age} = 47.75$ months, $SD = 7.37$) was to test a bifactor model of aggression, which allows for aggression to be assessed dimensionally, and postulates a co-occurring aggression factor as well as unique relational and physical aggression factors, during early childhood. Aggression was measured using reliable observer and teacher reports of physical and relational aggression subscales. The two-factor model was an adequate fit to the data but the bifactor model was a significant improvement in model fit compared to the two-factor model. Alternative statistics for evaluating bifactor models were used in the current study. The measurement invariance (e.g., configural invariance, metric invariance, and scalar invariance) of the bifactor model was tested across gender and results from the bifactor model using teacher report showed that the model was invariant across gender. Lastly, the criterion validity of the model was examined by evaluating the relations between the bifactor model and observations of physical and relational aggression and results generally supported the validity of the bifactor model. Overall, results suggest that a bifactor model of aggression may be a useful method for studying aggression in early childhood.

Keywords Relational aggression · Physical aggression · Early childhood · Bifactor model · Gender

Introduction

Historically, aggression was considered a unidimensional construct defined by physical acts such as hitting or kicking. Research in the field of aggression has established that aggression is actually a multidimensional construct, defined by both the form of behavior (i.e., physical or relational aggression) and the function of the behavior (i.e., proactive or reactive aggression; Crick and Grotpeter 1995; Dodge et al. 2006). Recent work has focused on elucidating the most effective methods for studying aggression in a multidimensional manner, particularly among the forms of aggression. Physical aggression is the intent to harm or control others through physical acts, such as hitting and kicking, and relational aggression is the intent to harm or control others through relationships, such as spreading malicious gossip, excluding others from play, and threats to end the relationship (Crick and Grotpeter 1995; Crick and Zahn-Waxler 2003; Dodge et al. 2006). This paper strives to evaluate a bifactor model of physical and relational aggression that is inclusive of co-occurrence among the forms of aggression as well as the unique components of physical and relational aggression in early childhood.

Early childhood is a unique time for aggression development. Research on trajectories of aggression suggests that physical aggression is more prevalent during this developmental period than later developmental periods (Côté et al. 2006; NICHD Early Child Care and Research Network, ECCRN 2004). Additionally, within early childhood, relational aggression becomes observable at around two and a half years of age (Crick et al. 2006a) and is fairly unsophisticated and often an immediate response to a problem (Crick et al.

✉ Kristin J. Perry
kperry5@buffalo.edu

1 Department of Psychology, University at Buffalo, The State University of New York, 478 Park Hall, Buffalo, NY 14260-4110, USA

Published online: 23 August 2017

2004). Additionally, relational aggression is direct and overt at this developmental period, where behaviors such as gossip will often occur right in front of the victim (Crick et al. 2004). The relatively high occurrence of physical aggression relative to other developmental periods and the emergence of relational aggression make the study of the co-occurrence of aggression important in early childhood.

In general, co-occurrence among physical and relational aggression is the norm rather than the exception. In general, the association among relational and physical aggression is high (*r* ranges from .50 to .70; Little et al. 2003). Despite the high association between the two constructs, they are differentially related to adjustment problems, such as internalizing and externalizing problems (Murray-Close et al. 2016).

Different statistical approaches have been used to take into account the overlap between the two forms of aggression. The most common approach involves controlling for one form of aggression when examining effects of the other form of aggression (i.e., controlling for physical aggression when relational aggression is the variable of interest; Smith et al. 2010). By using this approach, researchers can test the effects of one form of aggression while partialling out the variance in the other form of aggression (Smith et al. 2010). However, when interpreting results with this technique, researchers must consider what an aggression variable represents when you partial out the other aggression variable, considering the two are often highly correlated. Partialling out one form of aggression, fundamentally changes the aggression variable of interest (Miller and Chapman 2001). Additionally, this technique does not take into account the co-occurrence among the aggression variables, as it focuses on the unique contributions of each form of aggression. With the high correlations among the forms of aggression, co-occurrence seems to be the norm rather than the exception and therefore, researchers may benefit from taking into account co-occurrence.

When researchers have taken into account co-occurrence, two approaches have generally been used. The first approach is the cutoff-approach, where a child is arbitrarily classified as comorbidly aggressive if they are higher than one standard deviation (or ½ *SD* in some cases) above the mean for both relational and physical aggression (see Crick et al. 2006b). Research using this approach results in classifying children into four categories: (a) non-aggressive, where the child is not high on either relational or physical aggression; (b) relationally aggressive, where the child is only high on relational aggression but not physical aggression; (c) physically aggressive, where the child is only high on physical aggression and not relational aggression; (d) co-morbid aggressive, where the child is high on both physical and relational aggression (Crick et al. 2006b). This approach can be useful in clinical settings to assess which children may be at risk for a clinical level of aggressive behavior or associated psychopathology as it is a highly person centered approach (Smith et al. 2010). However, in research settings, categorizing children into aggression categories restricts the range of the aggression variables, potentially reducing the power to detect effects and resulting in a loss of measurement reliability (MacCallum et al. 2002). It also results in less information about the child's aggressive behavior. For example, a child who is 1.2 standard deviations above the mean in relational and physical aggression is in the same category as a child who is 3.0 standard deviations above the mean, and thus the category membership still does not convey information about the severity of the aggressive behavior. The second approach uses data-driven statistical analyses such as latent class analysis and growth mixture modeling to identify whether a child is in a low, moderate, or high relational and physical aggression class and then combines these to form a comorbid class (i.e., Moderate RA, High PA; Low RA, High PA; see Ettekal and Ladd 2015). This approach allows for more variability in class membership, as there are a greater number of classes than in a cutoff approach. Additionally, the approach is data driven so participants' scores rather than an arbitrary cutoff score determine class membership. However, when using this approach there is still a restriction of range because class analysis does not allow for testing of within group heterogeneity (Carragher and McWilliams 2011), as within a moderate RA group, there will still be some variability in RA scores and therefore, researchers still may be reducing power to detect effects.

Methods for Assessing Co-Occurrence

Bifactor models allow for a hierarchical model of constructs, such that they assess whether the indicators contribute to specific factors over and above their contribution to the general factor (Chen et al. 2012). These models are particularly well suited for measuring multidimensional constructs, as the common variance in the items is partialled into a general factor and the unique variance in the items is partialled into specific "pure" factors (Chen et al. 2012). Bifactor models have been used to assess whether relational aggression fits on the antisocial behavior construct in children (Tackett et al. 2013) and in adults (Krueger et al. 2007). Additionally, bifactor models have been used to examine the structure of specific disorders, such as ADHD in children and adolescents ages 6- to 18-years-old, where a bifactor model, with two or three specific factors comprising hyperactivity, impulsivity, and inattention symptoms, and a general factor was a good fit to the data, although most reliable variance in the specific factors was explained by the general factor (Rodenacker et al. 2016). A bifactor model is used in this study to model co-occurrence of aggression, where specific relational and physical aggression factors and a general co-occurring aggression factor are hypothesized to be a better fitting model than a two-factor relational and physical aggression model (see Fig. 1). Second order models have also been used to study multidimensional constructs with a common

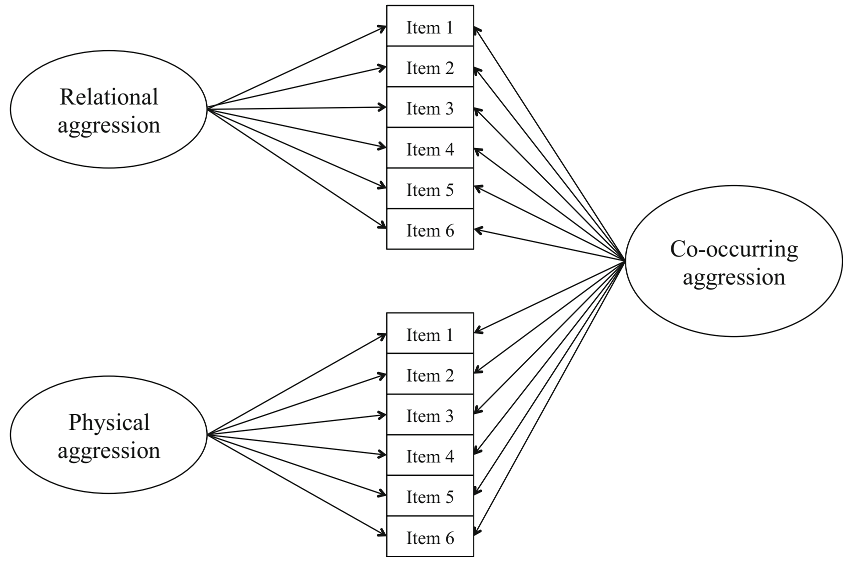

Fig. 1 A bifactor model of aggression using the relational and physical aggression subscales of the Preschool Social Behavior Scale-Observer report (PSBS-OR) and the Preschool Social Behavior Scale-Teacher form (PSBS-TF)

higher order factor (Chen et al. 2012). However, bifactor models are more appropriate than second order models in cases where a researcher is interested in the specific factors and wants to examine whether a specific factor exists over and above a general factor (Chen et al. 2012). In the current study, over and above general aggression, the specific relational and physical aggression facets of aggression are of a particular interest because one of the goals of the study is to examine whether aggression is in fact a multidimensional construct. Therefore, a bifactor model is better suited to achieve the goals of this study than a second order factor.

It is important to note that there have been important statistical and conceptual criticisms of bifactor models. At a conceptual level researchers have argued that bifactor models can be problematic because in real life there is no partitioning of behaviors to fit a statistical model (Rodriguez et al. 2016a). This is a valid point and speaks to the importance of using criterion data to assess the validity of the model. Second, there is a tendency for 'overfitting' when comparing bifactor models to other models where bifactor models may capture unwanted noise (Bonifay et al. 2016). To account for this overfitting, Bonifay et al. (2016) suggest using alternative statistics to evaluate bifactor models (Bonifay et al. 2016). These statistics include omega and omega hierarchical which are used to evaluate the reliability of unit weighted composites of the specific and general factors, explained common variance, and percentage of uncontaminated correlation which help determine whether the data is unidimensional or multidimensional, and construct replicability (H), which is particularly important for assessing the replicability of the loadings in a SEM framework (Rodriguez et al. 2016a, b). In a meta-analysis of bifactor models, Rodriguez et al. (2016a), found that the majority of prior data using bifactor models could be fitted as unidimensional without bias even though a bifactor model was found to be the best fit to these data in their individual studies (Rodriguez et al. 2016a). These alternate statistics will be used to evaluate the utility of the bifactor model in the current study.

Additionally, because of the criticisms of the utility and validity of the bifactor model, it is important to examine the criterion validity of the model. In the current study examining the relations between the model and observations of physical and relational aggression will test the criterion validity of the bifactor model. Observations of aggression are particularly useful for examining the validity of the model because they are a measure of observable behavior. The co-occurring factor may be related to both observations of physical and relational aggression.

Gender

Research has long supported the hypothesis that physical aggression is more prevalent among boys than girls in early childhood (Ostrov and Keating 2004; Ostrov et al. 2014). Some researchers have argued that girls are equally as aggressive as boys, but the type of aggression they display may be different (Crick and Grotpeter 1995; Crick and Zahn-Waxler 2003). This seems to be supported to a point. Relational aggression is the modal form of aggression among girls, such that girls are more likely to display relational compared to physical aggression (e.g., Ostrov et al. 2014). However, there may not be between group differences between boys' and girls' relational aggression such that a meta analysis found that boys are just as aggressive as girls when measuring indirect aggression, a construct that is highly related yet distinct from relational aggression (Card et al. 2008). Notably, a recent paper that collected data from nine countries found that boys are more physically aggressive than girls but there was no difference in relational aggression (Lansford et al. 2012). The research in this area is mixed, as method seems to play a role in whether researchers find a gender difference in relational aggression, with parent and teacher reports of aggression more likely to lead to a gender effect (Card et al. 2008). Additionally, even when using one informant, certain measurement instruments show gender differences whereas others do not. Results from a study of children in middle childhood found that for parent ratings of relational aggression there were gender differences when using the Children's Social Behavior Scale, but no gender differences when using the Social Relations Questionnaire (Tackett and Ostrov 2010). There is also a stronger association between indirect and direct aggression, two related but distinct constructs to relational and physical aggression, for boys compared to girls (Card et al. 2008), suggesting that girls are more likely to display one form of aggression without the other. The nature of the aggression also typically varies between girls and boys, with boys more likely to exhibit overt displays of aggression (i.e., hitting, directly excluding peer), and girls more likely to exhibit covert displays of aggression (i.e., threats of physical aggression, gossip) possibly because social norms discourage overtly aggressive acts from females (Lightdale and Prentice 1994). The gender non-normative theory of aggression postulates that children who display the gender non-normative form of aggression (i.e., physical aggression for girls, relational aggression for boys) have a greater negative effect on the child's outcomes (Crick 1997). These studies suggest that the structure of aggression may vary among boys and girls both in quantity (i.e., physical aggression is lower among girls) and quality (i.e., girls are more likely to display covert methods of aggression). Therefore, the invariance of the model will be assessed across gender.

This study has multiple aims: (a) evaluate the two-factor structure of the relational and physical aggression subscales of the Preschool Social Behavior Scale using an observer report (PSBS-OR), (b) compare a bifactor model of the PSBS-OR, which accounts for co-occurrence among physical and relational aggression and the unique or "pure" aspects of physical and relational aggression, to the two-factor model of the PSBS-OR, (c) assess whether the bifactor model is invariant across gender, (d) further validate a bifactor model of physical and relational aggression by testing the model using the PSBS teacher form (PSBS-TF), and (e) use criterion correlates of observations of aggression to support the utility of the bifactor model. Based on prior research it is hypothesized that the bifactor model will be a better fit to the data than a two-factor model because the bifactor model will account for the overlap between relational and physical aggression, in addition to the unique facets of each form of aggression and therefore, will explain more variability in the PSBS-OR measure. There is a paucity of research testing gender differences in a bifactor model of aggression, so no hypothesis is made about the invariance of gender. It is hypothesized that observations of the forms of aggression (i.e., physical and relational) will be related to both the co-occurring factor and the form specific factor (i.e., the relational aggression specific factor will be related to observations of relational aggression).

Method

Participants

This study used 164 participants (50.9% boys, *M* age = 47.75 months, *SD* = 7.37) from two separate independent samples conducted over a two-year period. Children were from relatively diverse backgrounds (4.9% African American, 9.9% Asian/ Pacific Islander/ Indian, 67.9% White, 3.7% Hispanic/Latino, 13.6% multiracial). Parental occupation was gathered at enrollment and was coded using Hollingshead's (1975) four-factor index 9-point scoring system (i.e., 9 = executives and professionals, 1 = service workers). Parents had the opportunity to enter two occupations per family, in which case the higher occupation code was taken. Parents' education was not taken and thus was not included in the total factor score. Values ranged from 2 to 9 with a 7.00 average, indicating that a typical family in our sample was from the third highest occupation group (i.e., 7 = small business owners, farm owners, managers, minor professionals), which suggests our sample is on average, middle class. Children were recruited from nine National Association for the Education of Young Children (NAEYC) accredited or recently accredited early childhood education centers in a large northeastern US city (16 classrooms). Four of the schools were university affiliated and five were community based. The county from which the study data was collected has a predominately white population, (13.5% African American, 3.2% Asian/ Pacific Islander/ Indian, 80.0% White, 1.8% multiracial, 1.5% race not listed) and a median income of $51,247 (U.S. Census Bureau 2010). Therefore, our study demographics are comparable to the larger county from which the samples were drawn.

The first sample used data from the control group of T1 of an intervention study, prior to randomization and intervention implementation (n = 61, *M* age = 48.81 months, *SD* = 7.01, 55.7% girls). Data for the second sample was collected a year after the data collection for the first sample with children who were not in the intervention sample (n = 105, *M* age = 47.37 months, *SD* = 7.56, 47.6% girls). Data for both samples were collected in the fall of the school year and were completely independent such that no children in the first sample were in the second sample. There were no differences between the two samples for ethnicity [χ^2 (6) = 2.73, p = .84, Cramér's V = .05], occupation [χ^2 (6) = 8.13, p = .23, Cramér's V = .09], gender [χ^2 (1) = 0.91, p = .34, Cramér's V = .07], or age [t(149) = 1.15, p = .25, Cohen's d = .19]. There were not expected crossover effects from the intervention into sample two because of the time lag between the samples (a year) and the independence of the samples. However, to determine whether the intervention produced any crossover effects into the second sample, differences in aggression scores were examined between the two samples. There were no differences between the two samples on relational aggression scores for teacher report [$t(160) = -1.98$, $p = .05$, Cohen's d = .31] or observer report [$t(160) = 0.42$, $p = .68$, Cohen's d = .06] or physical aggression scores for teacher report [$t(160) = -1.47$, $p = .14$, Cohen's d = .23] or observer report [$t(160) = -0.45$, $p = .65$, Cohen's d = .07]. Two children had missing data on all variables used in the study and therefore, were excluded from the analyses for a complete sample size of 164.

Procedure

All children in participating classrooms were invited to participate and parents provided written consent for their children's participation prior to beginning of the studies. Head teachers also provided written consent prior to report completion. This study was approved by the local institutional review board (IRB). Teachers were compensated \$10 - \$25 depending on their class size.

Measures

Observations of Aggression Initially, trained undergraduate and graduate research assistants collected naturalistic observations using a focal child sampling with continuous recording procedure as outlined in the Early Childhood Observation System (Ostrov and Keating 2004). There were a total of 17 trained undergraduate observers and three graduate observers. Prior to classroom entry, observers underwent stringent training by completing readings, discussing behavior via videotape, completing six standard observation sessions using videotape, and passing a written vignette test assessing their knowledge of the constructs. Observers were trained to identify relational and physical aggression and victimization, in addition to prosocial behavior (Ostrov and Keating 2004). Typically there were two to three observers per classroom. Observations were undertaken in a two-month period, with the goal of completing eight, ten-minute observation sessions per child. The number of physical and relational aggression acts was summed across the eight sessions to get a total relational and physical aggression score for each child. On average, each child had a total of 7.57, ten minutes observations at the end of the two-month period (i.e., approximately 76 min of observation per child) and observers spent approximately nine hours a week or 72 h total in the classroom. Observers spent a minimum of two days in the classroom prior to beginning observations to reduce reactivity to their presence and to conduct a live practice reliability session with the trainer. The average reactivity (i.e., child looking at observer, talking to observer, or talking about observer) was a 3.72 over the total eight sessions, which suggests that children were minimally reactive or not reactive at all to the observers. Reliability sessions were collected for 15% of observations and these sessions demonstrated that observations of physical and relational aggression were at an acceptable level of reliability (Relational aggression ICC = .76, Physical aggression ICC = .85).

Observer Reports of Aggression Observer ratings of behavior were used for reports of the PSBS-OR. Due to the large amount of time (i.e., 72 h over a two month period) observers spent in the classroom, observers are likely reliable informants of the child's behavior. After completing behavioral observations, one undergraduate observer from each classroom was randomly selected to complete reports for each participant that comprise the observer ratings of behavior, which are used in the present study. The timing of observer reports was standardized such that all observer reports were completed after observations were completed at the end of the two-month period. The undergraduate observers were not privy to the study hypotheses or predictions. Prior research has validated the use of observer reports and their high associations with teacher reports and significant overlap with naturalistic observations (e.g., Murray-Close and Ostrov 2009). Observers may be less biased than teachers because they are not explicitly interacting with children, and thus are not influenced by prior history or a relationship with the child (Ostrov and Hart 2013). Additionally, observers are trained to systematically observe aggressive behavior, and thus may be more valid informants of aggression.

Preschool Social Behavior Scale-Observer Report and Teacher Report Physical and relational aggression was measured with the Preschool Social Behavior Scale-Observer Report (PSBS-OR) and Preschool Social Behavior Scale-Teacher Form (PSBS-TF, Crick et al. 1997; Ostrov 2008). In the current study, the physical and relational aggression subscales were used and each contained six questions rated on a 5-point Likert scale (1 = Never or almost never true; 5 = Always or almost always true). Originally, the PSBS was designed as a teacher report measure, where a principal-components analysis confirmed four factors: physical aggression, relational aggression, prosocial behavior, and depressed affect (Crick et al. 1997). Importantly, the original principal components analysis confirmed the presence of two factors among the relational and physical aggression items, suggesting delineation between the two constructs (Crick et al. 1997). Observer reports have previously been highly correlated with teacher reports of the PSBS (Ostrov 2008). In the current sample, the PSBS-OR was associated with the PSBS-TF for the relational aggression subscale (r = .30, p < .001) and the physical aggression subscale (r = .53, p < .001), and correlated with observations of aggression for the relational aggression subscale (r = .40, p < .001) and the physical aggression subscale (r = .60, p < .001). The subscales have demonstrated acceptable reliability in prior preschool samples (e.g., Ostrov 2008) and had good internal consistency in this sample (Cronbach's rs > .90). One advantage of the PSBS-OR compared to the PSBS-TF is that observers are specifically trained to identify aggressive behavior and therefore, their reports of aggression on the PSBS may be more accurate.

Data Analysis Plan

First, descriptive data of the measures were obtained, including means, standard deviations, and an analysis of outliers. An outlier was defined as any value that is greater than three standard deviations above or below the mean. Outliers were modified by adjusting the outlier value to +/− three standard deviations from the mean (Kline 2011). Skew statistics were assessed, where skew values ranged from .45 to 1.88 and kurtosis statistics ranged from −.99 to 2.90, which suggests that any non-normality in the data did not impact the analyses (Kline 2011). Additionally, because students are nested within classrooms, the intraclass correlation coefficient (ICC) was examined for the PSBS observer and teacher report to assess whether there was a clustering effect of classroom on the key variables in this study. Results showed that for observer report there was no between group variance in classrooms for physical aggression (ICC = .005) or relational aggression (ICC = .007). For teacher report there was between group variance in classrooms for physical aggression (ICC = .26) and relational aggression (ICC = .23). This would make multilevel modeling techniques ideal for the teacher reported models. However, past researchers have suggested that 30 groups be the minimum number of groups needed to use multilevel modeling techniques (Kreft and De Leeuw 1998), which is much higher than the 15 classrooms we have in the current study. Therefore, to control for dependency in the data, a classroom membership variable was created which identified the classroom the child was enrolled in. This variable was entered as a control variable in the teacher reported models.

Maximum likelihood estimation with robust standard errors (MLR) has been shown to outperform other estimation methods designed for ordinal data when small samples are used and when scales have five or more categories (Li 2016; Rhemtulla et al. 2012**)**. Therefore, MLR was used in the current study due to the small sample size and the nature of the data. This study uses data from one time point and therefore, missing data was minimal (< 1% of the data was missing) and was accommodated by using full information maximum likelihood (FIML).

All models were estimated in Mplus version 7.4 (Muthén and Muthén 1998-2015). The likelihood ratio χ^2 test was used to test overall model fit where p > .05 indicates good model fit. Alternative fit indices were also used to determine model fit because the chi-squared test has been found to be an unreliable and restrictive indicator of model fit (Hooper et al. 2008). The comparative fit index (CFI), where values greater than .90 suggest adequate fit and values greater than .95 suggest good fit, the standardized root mean-square residual (SRMR) fit index where values less than .08 represent adequate model fit and values less than .05 represent good model fit (Hu and Bentler 1999), and the root mean square error of approximation (RMSEA; Steiger 1990), where values greater than .10 represent poor fit, values less than .08 represent mediocre fit, and values less than .05 represent close fit (Browne and Cudeck 1992; MacCallum et al. 1996) were considered. To test comparisons in model fit (i.e., testing the two-factor model compared to the bifactor model) and when testing measurement invariance, chi-square difference tests were used. MLR uses a scaling factor to adjust the chi-square test statistic and therefore, direct comparison of the chi-square test statistic cannot be used. Methods developed by Satorra and Bentler (2010) were used to calculate the chi-square difference test statistic. Additionally, when testing differences in fit when evaluating the measurement invariance of the model, ΔCFI was used in addition to chi-square difference tests. The ΔCFI test, where ΔCFI > .01 indicates a significant difference in model fit, has been used to test the measurement invariance of previous bifactor models (e.g., Ebesutani et al., 2014) and has been found to be less susceptible to the effects of sample size and model complexity (Chen, 2007). Therefore, in the current study we included chi-square test difference statistics and Δ CFI statistics for testing differences in model fit when testing for measurement invariance. If either the chi-square difference test or the Δ CFI statistic indicated no difference in model fit, then the more parsimonious model was considered the better fitting model. If the chi-square difference test indicated no difference in model fit, then the more parsimonious model was considered the better fitting model.

Given that the PSBS has not previously been validated with an observer as the informant, a two-factor model of physical and relational aggression was first estimated. Next, the bifactor model was estimated and compared to the two-factor model. Last, the invariance of the bifactor model was tested across gender according to procedures outlined by Schmitt and Kuljanin (2008). Consistent with these procedures, to test the measurement invariance of the best fitting model across gender, first a configural invariance model is tested where the bifactor model is estimated separately for males and females and model fit is examined. Second, a metric invariance model is tested where the factor loadings are held to equivalency across males and females and compared to the configural model.

Third, a scalar invariance model is tested where the factor loadings and intercepts are held to equivalency across gender.

In the case of noninvariance, partial invariance was assessed using sequential use of modification indices (MI) under the full metric invariance model to determine which parameters should be sequentially freed in accordance with procedures outlined by Yoon and Millsap (2007). First the parameter with the largest MI was freed where the fit indices for this model were compared to the configural invariance model. If there were still differences between the configural invariance model and the new model then MI were examined and the parameter with the largest MI statistic was freed, while also freeing the first parameter, which was compared to the configural invariance model. This process continued until there was no difference between the configural invariance model and the new model.

To evaluate the criterion validity of the model, two path analytic models were tested using observations of physical and relational aggression at the same time point. Observations of physical and relational aggression were regressed on the bifactor model, allowing for an association between the two subtypes of aggression.

To further validate the use of the bifactor model using the PSBS-OR, the same steps were tested using the PSBS-TF. Using the PSBS-TF, first a two-factor model was tested, followed by a bifactor model, the invariance of the bifactor model across gender was tested, and the criterion validity of the model was assessed using observations of aggression. In prior work using this observation scheme and the PSBS, teacher reports of relational aggression have been correlated with relational aggression observations in the magnitude of .13 to .55 and teacher reports of physical aggression have been correlated with physical aggression observations in the magnitude of .13 to .56 (e.g., Crick et al. 2006a; Ostrov et al. 2008, 2013). Therefore, we expect the criterion validity statistics to be in this range of magnitude.

Alternate Statistics

In addition to the traditional methods for evaluating the bifactor model, such as examining fit indices and comparing the model to a two factor model, we will be using coefficient omega, coefficient omega hierarchical (Omega H), explained common variance (ECV), percentage of uncontaminated correlation (PUC), and construct replicability (H) to evaluate model fit. Omega is a reliability estimate of the general and specific subscales for factor analytic models, using all sources of common variance (Rodriguez et al. 2016b). In contrast to Omega, Omega H, uses only the variance in scores that can be attributable to a certain factor instead of all of the common variance (Rodriguez et al. 2016b). Omega H is increasingly important to consider for the specific factors because using the common variance may lead to the subscale scores looking reliable, when they in fact are not (Rodriguez et al. 2016b). For subscale scores, Omega H is calculated by using the specific variance attributable to the specific factor while controlling for the general factor variance (Rodriguez et al. 2016b). Omega and Omega H are useful for determining whether the subscales are reliable, how much reliable variance is attributable to general vs. specific factors, and whether researchers can use unit-weighted scores (Rodriguez et al. 2016b).

In an SEM framework, it is important to assess construct replicability, which measures whether a group of items represent a latent factor well (Rodriguez et al. 2016b). H is used in a SEM framework to assess the potential for model specification using an item set (Rodriguez et al. 2016b). When H is high, there is a greater likelihood of being able to replicate the model in subsequent studies because the latent factor is well defined by the items (Rodriguez et al. 2016b). Past researchers have set a standard value of H at .70 to ensure that a factor is represented adequately (Hancock and Mueller 2001).

Explained common variance (ECV) and Percentage of Uncontaminated Correlation (PUC) are used in conjunction to help determine whether a model is actually multidimensional (Rodriguez et al. 2016b). Researchers must move beyond fit indices supporting the fit of a bifactor model to the data, as a bifactor model is susceptible to overfitting, such that the data may be 'essentially unidimensional' despite fit indices suggesting a bifactor model is a superior fit. ECV evaluates the strength of the general factor compared to the specific factors, by dividing the variance explained by the general factor by the variance explained by all factors (Rodriguez et al. 2016b). ECV varies from Omega H because Omega H focuses on the variance explained in a unit-weighted composite (Rodriguez et al. 2016b). PUC orients ECV to the structure of the data and PUC and ECV used together can demonstrate whether a model is biased by fitting multidimensional data to a unidmensional model (Rodriguez et al. 2016b). PUC evaluates how many correlations in the data are used on the general factor (Rodriguez et al. 2016b). Values from Rodriguez et al. (2016a) suggest that when ECV > .70 and PUC > .70 there is little bias in SEM when fitting a multidimensional model in a unidimensional manner.

Results

Two-Factor Model Vs. Bifactor Model

A two-factor model of physical and relational aggression of the PSBS-OR was an adequate fit to the data [$\chi^2(53) = 134.21$, $p < .001$, $CFI = .93$, $SRMR = .05$, $RMSEA = .097$]. All standardized factor loadings were significant at $p < .001$ and values of the factor loadings ranged from .74 to .91 for the relational aggression factor and .77 to .90 for the physical aggression factor. There was a significant association between physical and relational aggression (r = .63, p < .001). These results are consistent with the factor structure of the PSBS found in Crick et al. (1997) initial work with the measure suggesting that relational and physical aggression are related, but separate factors.

The bifactor model of the PSBS-OR was an adequate fit to the data [$\chi^2(42) = 91.54$, $p < .001$, $CFI = .96$, $SRMR = .04$, $RMSEA = .09$] and was a significantly better fit than the two-factor model [$\Delta\chi^2(11) = 37.70$, $p < .001$]. The standardized factor loadings are shown in Table 1. The association between the two specific factors, relational and physical aggression, was constrained to zero, which is consistent with a traditional bifactor model (Chen et al. 2012). Three of the physical aggression items did not significantly load on the specific factor. This indicates that more of the variance in these physical aggression items is being partialled into the co-occurring factor. The physical aggression items in general loaded more strongly on the co-occurrence factor than the relational aggression items, suggesting that it may be more likely to be relationally aggressive without being physically aggressive than it is to be physically aggressive without being relationally aggressive.

Gender Invariance

The invariance of the bifactor model of the PSBS-OR was tested across gender. First, the model was estimated separately in each group. The bifactor model provided an adequate fit to the data for girls [$\chi^2(42) = 59.07$, $p = .04$, $CFI = 0.97$, $SRMR = .04$, $RMSEA = .07$]. However, the model was a poor fit to the data for boys [$\chi^2(42) = 113.39$, $p < .001$, $CFI = 0.89$, $SRMR = .07$, RMSEA = .14]. In the present case, where the model is a poor fit for boys, it is not appropriate to test the measurement invariance of the model (Muthén 2015) and therefore, we tested the criterion validity of the model in the entire sample.

Criterion Validity

A path analytic model was used to test the relations between the bifactor model and observations of aggression, where observations of physical and relational aggression were regressed on the bifactor model. The model was a good fit to the data [$\chi^2(60) = 116.81$, $p < .001$, $CFI = 0.95$, $SRMR = .04$, $RMSEA = .08$]. Observations of physical and relational aggression were significantly associated (β = .28, p = .001). The co-occurring factor was positively predictive of physical aggression observations (β = .51, p < .001) and there was a trend towards the co-occurring factor being positively predictive of relational aggression observations (β = .16, p = .09). Additionally, there was specificity in the associations, such that the relational aggression factor was positively predictive of relational aggression observations (β = .41, p < .001) and not related to the physical aggression observations (β = −.06, p = .34). Similarly, the physical aggression factor was positively predictive of physical aggression observations (β = .34, p = .03) but was not related to relational aggression observations (β = .16, p = .16).

Table 1 Standardized factor loadings for the PSBS-OR bifactor model

PSBS-OR item number	Relational aggression	Physical aggression	Co-occurring aggression
6"*This child tells a peer that he/she won't play with that peer or be that peer's friend unless he/ she does what the child asks*"	.73		.56
12"*This child tells others not to play with or be a peers friend*"	.61		.57
15 "*When mad at a peer, this child keeps that peer from being in the play group*"	.68		.49
17 "*This child tells a peer they won't be invited to their birthday party unless he or she does what the child wants*"	.54		.51
22 "*This child tries to get other to dislike a peer* (e.g., *by whispering mean things about the peer,* etc.)"	.63		.54
24 "*This child verbally threatens to keep a peer out of the play group if the peer doesn't do what the child asks*"	.66		.57
4 "*This child kicks or hits others*"		.48	.81
8 "*This child verbally threatens to hit or beat up other children*"		−.13†	.94
11 "*This child pushes or shoves other children*"		.52	.71
16 "*This child verbally threatens to physically harm another peer in order to get what they want*"		.05†	.86
19 "*This child ruins other peer's things when he or she is upset*"		.29	.71
25 "*This child hurts other children by pinching them*"		.16†	.75

All factor loadings are significant (p < .05) unless indicated by a †

PSBS-OR Preschool Social Behavior Scale-Observer Report

Validation Using the PSBS-TF

Child classroom membership was controlled for in all teacher reported models. A two-factor model of physical and relational aggression provided an adequate fit to the data [$\chi^2(63) = 150.78$, $p < .001$, $CFI = .91$, $SRMR = .06$, $RMSEA = .09$]. All standardized factor loadings were significant at $p < .001$ and values of the factor loadings ranged from .77 to .85 for the relational aggression factor and .84 to .89 for the physical aggression factor. There was a significant association between physical and relational aggression ($r = .76$, $p < .001$). Similar to the PSBS-OR, the bifactor model of the PSBS-TF provided a good fit to the data [$\chi^2(51) = 56.64$, $p = .27$, $CFI = .99$, $SRMR = .03$, $RMSEA = .03$] and was a significantly better fit than the two-factor model [$\Delta\chi^2(12) = 77.89$, $p < .001$]. The standardized factor loadings are shown in Table 2.

The measurement invariance of the bifactor model of the PSBS-TF was tested across gender. The configural invariance model entailed a multiple group CFA in which the same model was estimated freely in both groups. The bifactor model provided a good fit to the data for girls [$\chi^2(51) = 63.46$, $p = .11$, $CFI = .98$, $SRMR = .05$, $RMSEA = .06$] and a good fit to the data for boys [$\chi^2(51) = 59.65$, $p = .19$, $CFI = 0.98$, $SRMR = .04$, $RMSEA = .05$]. A configural invariance model was tested in which the model was estimated separately for both groups, which provided a good fit to the data [$\chi^2(102) = 121.94$, $p = .09$, $CFI = .98$, $SRMR = .04$, $RMSEA = .05$]. Next, a metric invariance model was tested that constrained the factor loadings to equivalency across groups. The metric invariance model provided an adequate fit to the data [$\chi^2(126) = 151.10$, $p = .06$, $CFI = .98$, $SRMR = .11$, $RMSEA = .05$] and there was no difference in model fit compared to the configural invariance model [$\Delta\chi^2(24) = 28.95$, $p = .22$, ΔCFI = .00].

Next a scalar invariance model was tested in which the factor loadings and intercepts were constrained to equivalency across gender. The model was an adequate fit to the data [$\chi^2(138) = 164.59$, $p = .06$, $CFI = .98$, $SRMR = .11$, $RMSEA = .05$] and there was no difference in model fit between the scalar invariance model and the configural invariance model [$\Delta\chi^2(36) = 42.78$, $p = .20$, ΔCFI = .00]. Therefore, the bifactor model using teacher report was considered invariant across gender.

A path analytic model was used to test the relations between the bifactor model and observations of aggression, where observations of physical and relational aggression were regressed on the bifactor model. The model was a good fit to the data [$\chi^2(69) = 83.87$, $p = .11$, $CFI = 0.99$, $SRMR = .04$, $RMSEA = .04$]. Physical and relational aggression observations were significantly associated ($\beta = .30$, $p < .001$). The co-occurring factor was positively predictive of physical aggression observations ($\beta = .35$, $p = .001$), the physical aggression factor was positively predictive of physical aggression observations ($\beta = .45$, $p = .001$), and none of the factors predicted relational aggression observations.

Table 2 Standardized factor loadings for the PSBS-TF bifactor model

PSBS-TF item number	Relational aggression	Physical aggression	Co-occurring aggression
6"*This child tells a peer that he/she won't play with that peer or be that peer's friend unless he/ she does what the child asks*"	.54		.56
12"*This child tells others not to play with or be a peers friend*"	.60		.64
15 "*When mad at a peer, this child keeps that peer from being in the play group*"	.59		.62
17 "*This child tells a peer they won't be invited to their birthday party unless he or she does what the child wants*"	.61		.57
22 "*This child tries to get others to dislike a peer* (e.g., *by whispering mean things about the peer,* etc.)"	.29		.74
24 "*This child verbally threatens to keep a peer out of the play group if the peer doesn't do what the child asks*"	.18†		.84
4 "*This child kicks or hits others*"		.67	.70
8 "*This child verbally threatens to hit or beat up other children*"		.14†	.89
11 "*This child pushes or shoves other children*"		.59	.65
16 "*This child verbally threatens to physically harm another peer in order to get what they want*"		.13†	.89
19 "*This child ruins other peer's things when he or she is upset*"		.43	.72
25 "*This child hurts other children by pinching them*"		.31	.79

All factor loadings are significant ($p < .05$) unless indicated by a †

PSBS-TF Preschool Social Behavior Scale-Teacher Form

Alternate Statistics

Alternate statistics, including Omega, Omega hierarchical, Omega hierarchical subscale, H an index of construct replicability, explained common variance (ECV), and percentage of uncontaminated correlation (PUC) were calculated for the two bifactor models and are shown in Table 3. Omega values ranged from .95 to .97 for all models. Omega H values for the general factor ranged from .76 to .82 and Omega H values for the subscales ranged from .03 to .18, suggesting that most of the reliable variance lies in the general factor. H values ranged from .45 to .95, and importantly across reporter the relational aggression and co-occurring factors had H values ranging from .69 to .95, which approach and meet the .70 threshold as set by Hancock and Mueller (2001). These statistics suggest that across reporter the items define the co-occurring and relational aggression latent factors well. The same was not true for the physical aggression factor which had H values of .45 and .63, suggesting that the physical aggression items do not define the specific physical aggression factor well after partialling out the variance in the items in the co-occurring factor. Lastly, multidimensionality was assessed using ECV and PUC, where values from ECV > .70 and PUC > .70 suggest that there is little bias by fitting the multidimensional data in a unidimensional way (Rodriguez et al. 2016a). When using observer reports there seems to be more evidence for multidimensionality, where ECV was .64 and PUC was .55 across informants. When using teacher reports, there is more evidence for the unidimensionality of aggression as there may be a small level of bias in fitting the bifactor model to a unidimensionl model because ECV is .71 and PUC is .55.

Table 3 Alternate statistics for the bifactor models

	Omega	Omega H	H	ECV
Bifactor- OR				.64
General	.96	.76	.95	
RA	.96	.18	.82	
PA	.95	.03	.45	
Bifactor- TF				.71
General	.97	.82	.95	
RA	.96	.09	.69	
PA	.96	.06	.63	
PUC for all scales = .55				

TR Teacher form, *OR* Observer report, *RA* Relational aggression, *PA* Physical aggression, *Omega H* Omega Hierarchical, *H* construct replicability, *ECV* Explained common variance, *PUC* Percentage of Uncontaminated Correlation (PUC)

Discussion

The goal of this preliminary study was to introduce a novel method for evaluating co-occurrence of physical and relational aggression, while assessing the invariance of the model across gender, and comparing the bifactor model to a two-factor model of relational and physical aggression in early childhood. Early childhood is a unique time for the development of aggression because aggression is still relatively overt, which allows for outside observers to be aware of the behavior, and children begin using both types of aggressive behavior. Results suggested that a two-factor model was a good fit to the data, which confirmed the factor structure of the physical and relational subscales of the PSBS (Crick et al. 1997) using a trained observer as the informant. Consistent with predictions, the bifactor model was a better fit to the data than the two-factor model, suggesting that accounting for co-occurrence among relational and physical aggression, over and above the specific aspects of the constructs, results in a more accurate model of aggressive behavior. The utility of the bifactor model over a two-factor model was also tested using teacher reports of the PSBS to add to the validity of the model. Results from the analyses using the PSBS-TF were consistent with the results found using the PSBS-OR, where a bifactor model was a better fit to the data than a two-factor model. Measurement invariance was evaluated and the bifactor model was found to be invariant across gender for teacher report. However, we were unable to examine the measurement invariance of the bifactor model when using observer report because the bifactor model was not a good fit for boys. Overall, results from this study suggest that a bifactor model may be a useful way to assess relational aggression, physical aggression, and their co-occurrence in early childhood although future replication is necessary to further understand the structure of aggression in this developmental period.

The bifactor model of aggression was a better fit to the data then the two-factor model in the current study. There are several advantages to bifactor models, such that they allow for studying the specific nature of relational and physical aggression after partialling out the common variance between the two (Chen et al. 2012). Prior researchers have discussed the advantages of a bifactor model when scholars are interested in both the specific factors and general factor (Chen et al. 2012). This may be particularly advantageous when studying aggression because the bifactor model allows for the unique characteristics of physical and relational aggression to be measured after accounting for the substantial overlap between the two constructs. Researchers can use this method in lieu of controlling for one form of aggression when assessing the other form of aggression, because the bifactor model allows researchers to examine the unique facets of physical and relational aggression and their links with psychosocial outcomes, while also accounting for their co-occurrence.

The current study utilized alternate statistics to evaluate the bifactor model. Overall, these statistics provided mixed evidence for aggression as a multidimensional construct that was well represented by the items in the current data set. There was more evidence for the multidimensionality of aggression when using observer reports of aggression (ECV = .64, PUC = .55) compared to teacher reports of aggression (ECV = .71, PUC = .55), although it is important to note that the ECV value of .64, which is used as evidence for multidimensionality of aggression, is not incredibly different from the value of .70, so there still may be relatively little bias in fitting the observer report model in a unidimensional manner. Results also suggested that most of the reliable variance in raw scores was due to the co-occurring aggression factor, which suggests that unit weighted composite scores may be unreliable in predicting outcomes. Across informant, construct replicability (H) was acceptable for the items on the specific relational aggression factor and the co-occurring aggression factor and was at unacceptable levels for the items on the specific physical aggression factor. After partialling out the variance in the items on the co-occurring factor, the physical aggression items may not represent the specific physical aggression factor well. Overall, results from the alternate statistics demonstrate the importance of replicating the model in different samples using different measures of aggression to determine whether aggression is a unidimensional or multidimensional construct.

Observations of aggression were used to evaluate the criterion validity of the bifactor model. In general, for observer and teacher reports, the bifactor model was related to observations in the way hypothesized and in the magnitude that is consistent with prior research examining correlations between teacher reports of aggression and observations of aggression. There was one exception. For teacher reports the bifactor model was not related to relational aggression observations. There are two potential explanations for this lack of effect. First, relational aggression is more difficult to identify and there is often less between rater agreement for relational aggression compared to physical aggression as evidenced by the lower inter rater agreement using the PSBS in the current study for relational compared to physical aggression. Second, the observations themselves are less reliable for relational aggression compared to physical aggression, as evidenced by the relatively lower ICCs for relational aggression compared to physical aggression. Future research should continue to develop unique and reliable methods for measuring physical and relational aggression.

Limitations

Despite the novelty of the work, there are limitations to the current study. The participants in this study were in early childhood and thus results may not be generalizable to other developmental periods. Similarly, participants from the study were from a limited number of schools from one area in the northeastern US, and thus, the results may not be generalizable to other geographic regions or cultures. Additionally, this study had a relatively small sample size for the SEM approach, underscoring the importance of future replication. However, with the unique nature of time consuming and intensive observations, which allowed us to evaluate the criterion validity of the model using actual behavior correlates, the sample size was fairly large. One important limitation of bifactor models is that they may have limited practical or applied use because in reality variability in behavior cannot be statistically partialled apart. Moreover, we echo a call for more research on verbal aggression (Ostrov and Kamper 2015) and emphasize the importance of expanding the present findings to other forms of aggression, such as verbal and non-verbal aggression. The current study used the same measure to examine aggression among the two informants, and used observations of aggression to validate the model when using observer reports of aggression. Therefore, the replication of the bifactor model across informant is limited by the use of the same measure and the test of the criterion validity of the observer report bifactor model with observations of aggression is limited because the observers had access to the observations. Additionally, due to the small sample size of the current study, we were unable to use multilevel modeling techniques to account for the clustering effect of students within classroom for the teacher reports of aggression. We controlled for classroom membership in the current study to try to account for some of this nested effect, but future research would benefit from the use of multilevel modeling which would reduce any confounding effect of classroom on the models.

Future Directions

Future research should first and foremost focus on replicating the model and testing the utility of a bifactor model and the invariance of the model across gender in other developmental periods. Early childhood is a unique age period for aggression compared to later periods as physical aggression is more prevalent, the on-set of relational aggression is still emerging, and aggressive behavior is relatively overt, and therefore, the structure of aggression may vary through developmental periods. Moreover, the nature of peer relationships and friendships change during this period of development and gender segregation patterns often become more entrenched as children enter middle childhood (Hartup and Stevens 1999; Maccoby 1990), which may impact the present gender effects. Second, researchers should focus on examining the predictive validity of the bifactor model in longitudinal work, particularly the utility of the specific factors over and above the co-occurring factor when predicting psycho-social outcomes in prospective longitudinal work. Lastly, it may be beneficial to evaluate a bifactor model that is inclusive of the functions (i.e., proactive and reactive) of aggression, in addition to other forms of aggression, such as verbal aggression.

Conclusion

This study evaluated the utility of a bifactor model of aggression by comparing the model to a two-factor model of aggression and testing the invariance of the model across gender for both teacher and observer informants of behavior. Consistent with prior literature from other developmental periods (Tackett et al. 2013), the bifactor model was a better fit to the data than the two-factor model and the model was invariant across gender when using teacher report. There was evidence for the criterion validity of the bifactor model, as the bifactor model was related to observations of physical and relational aggression in a theoretically consistent way. Overall, this study introduces a bifactor model of the forms of aggression in early childhood and demonstrates the potential utility of this model in examining aggression for children in early childhood.

Acknowledgements We thank the UB Social Development Lab for their assistance and special thanks to Dr. Kimberly Kamper-DeMarco and Sarah Blakely-McClure for methodological assistance and coordination of the data collection, Amanda Levy for reference checking, and Miriam Stotsky for APA and formatting consultation. We are grateful to the families, teachers, and administrators of participating schools.

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Funding Preparation of this manuscript was facilitated by a grant from the National Science Foundation to the second author (BCS-1450777).

Conflict of Interest The authors declare that they have no conflict of interest.

References

Bonifay, W., Lane, S. P., & Reise, S. P. (2016). Three concerns with applying a bifactor model as structure of psychopathology. *Clinical Psychological Science, 5*, 184–186. doi:10.1177/2167702616657069.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230–258.

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (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.

Carragher, N., & McWilliams, L. A. (2011). A latent class analysis of DSM-IV criteria for pathological gambling: results from the National Epidemiologic Survey on alcohol and related conditions. *Psychiatry Research, 187*, 185–192. doi:10.1016/j.psychres.2010.12.022.

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

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: a comparison of the bifactor model to other approaches. *Journal of Personality, 80*, 219–251. doi:10.1111/j.14676494.2011.00739.x.

Côté, S. M., Vaillancourt, T., LeBlanc, J. C., Nagin, D. S., & Tremblay, R. E. (2006). The development of physical aggression from toddlerhood to pre-adolescence: a nation wide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology, 34*, 71–85. doi:10.1007/s10802-005-9001-z.

Crick, N. R. (1997). Engagement in gender normative versus nonnormative forms of aggression: links to social psychological adjustment. *Developmental Psychology, 33*, 610–617. doi:10.1037/0012-1649.33.4.610.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66*, 710–722. doi:10.2307/1131945.

Crick, N. R., & Zahn-Waxler, C. (2003). The development of psychopathology in females and males: current progress and future challenges. *Development and Psychopathology, 15*, 719–742. doi:10.1017/S095457940300035X.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–588. doi:10.1037/0012-1649.33.4.579.

Crick, N. R., Ostrov, J. M., Appleyard, K., Jansen, E. A., & Casas, J. F. (2004). Relational aggression in early childhood: 'You can't come to my birthday party unless...'. In M. Putallaz & K. L. Bierman (Eds.), *Aggression, antisocial behavior, and violence among girls: A developmental perspective* (pp. 71–89). New York: Guilford Publications.

Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., & Ralston, P. (2006a). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology, 27*, 254–268. doi:10.1016/j.appdev.2006.02.006.

Crick, N. R., Ostrov, J. M., & Werner, N. E. (2006b). A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment. *Journal of Abnormal Child Psychology, 34*, 131–142. doi:10.1007/s10802-005-9009-4.

Dodge, K. A., Coie, J. D., & Lynam, D. (2006). Aggression and antisocial behavior in youth In W. Damon (series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3, Social, emotional, and personality development* (*6th ed.*, pp. 719-788). Hoboken: Wiley.

Ebesutani, C., McLeish, A. C., Luberto, C. M., Young, J., & Maack, D. J. (2014). A bifactor model of anxiety sensitivity: analysis of the anxiety sensitivity index-3. *Journal of Psychopathology and Behavioral Assessment, 36*, 452–464. doi:10.1007/s10862-013-9400-3.

Ettekal, I., & Ladd, G. W. (2015). Costs and benefits of children's physical and relational aggression trajectories on peer rejection, acceptance, and friendships: variations by aggression subtypes, gender, and age. *Developmental Psychology, 51*, 1756–1770. doi:10.1037/dev0000057.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and* *future—A Festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

Hartup, W. W., & Stevens, N. (1999). Friendships and adaptation across the life span. *Current Directions in Psychological Science, 8*, 76–79. doi:10.1111/1467-8721.00018.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*, 53–60.

Hu, L.-T., & Bentler, P. M. (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.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kreft, I. G. G., & De Leeuw, J. (1998). *Introducing multilevel modeling*. Newbury Park: Sage.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology, 116*, 645–666. doi:10.1037/0021-843X.116.4.645.

Lansford, J. E., Skinner, A. T., Sorbring, E., Giunta, L. D., Deater-Deckard, K., Dodge, K. A., et al. (2012). Boys' and girls' relational and physical aggression in nine countries. *Aggressive Behavior, 38*(298), 308. doi:10.1002/ab.21433.

Li, C. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936–949. doi:10.3758/s13428-015-0619-7.

Lightdale, J. R., & Prentice, D. A. (1994). Rethinking sex differences in aggression: aggressive behavior in the absence of social roles. *Personality and Social Psychology Bulletin, 20*, 34–44. doi:10.1177/0146167294201003.

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

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149. doi:10.1037/1082-989X.1.2.130.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40. doi:10.1037/1082-989X.7.1.19.

Maccoby, E. E. (1990). Gender and relationships: a developmental account. *American Psychologist, 45*, 513–520. doi:10.1037/0003-066X.45.4.513.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110*, 40–48. doi:10.1037/0021-843X.110.1.40.

Murray-Close, D., & Ostrov, J. M. (2009). A longitudinal study of forms and functions of aggressive behavior in early childhood. *Child Development, 80*, 828–842. doi:10.1111/j.1467-8624.2009.01300.x.

Murray-Close, D., Nelson, D. A., Ostrov, J. M., Casas, J. F., & Crick, N. R. (2016). Relational aggression: A developmental psychopathology perspective. In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd ed., pp. 660–722). New York: Wiley.

Muthén, L. K. (2015). Re: testing for measurement and structural equivalence [Online discussion group]. Retrieved from: http://www.statmodel.com/discussion/messages/11/21028.html?1429024985.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus (version 7.4)*. Los Angeles: Muthén & Muthén.

NICHD Early Child Care and Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: III. Person-centered trajectories of physical aggression. *Monographs of the Society for Research in Child Development, 69*, 41–49.

Ostrov, J. M. (2008). Forms of aggression and peer victimization during early childhood: a short-term longitudinal study. *Journal of Abnormal Child Psychology, 36*, 311–322. doi:10.1007/s10802-007-9179-3.

Ostrov, J. M., & Hart, E. J. (2013). Observational methods. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods (vol. 1): Foundations* (pp. 286–304). New York: Oxford University Press.

Ostrov, J. M., & Kamper, K. E. (2015). Future directions for research on the development of relational and physical peer victimization. *Journal of Clinical Child and Adolescent Psychology, 44*, 509–519. doi:10.1080/15374416.2015.1012723.

Ostrov, J. M., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: an observational study. *Social Development, 13*, 255–277. doi:10.1111/j.14679507.2004.000266.x.

Ostrov, J. M., Ries, E. E., Stauffacher, K., Godleski, S. A., & Mullins, A. D. (2008). Relational aggression, physical aggression and deception during early childhood: a multimethod, multi-informant short-term longitudinal study. *Journal of Clinical Child and Adolescent Psychology, 37*, 664–675. doi:10.1080/15374410802148137.

Ostrov, J. M., Gentile, D. A., & Mullins, A. D. (2013). Evaluating the effect of educational media exposure on aggression in early childhood. *Journal of Applied Developmental Psychology, 34*, 38–44. doi:10.1016/j.appdev.2012.09.005.

Ostrov, J. M., Kamper, K. E., Hart, E. J., Godleski, S. A., & Blakely-McClure, S. J. (2014). A gender-balanced approach to the study of peer victimization and aggression subtypes in early childhood. *Development and Psychopathology, 26*, 575–587. doi:10.1017/S0954579414000248.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*, 354–373. doi:10.1037/a0029315.

Rodenacker, K., Hautmann, C., Görtz-Dorten, A., & Döpfner, M. (2016). The factor structure of adhd – Different models, analyses and informants in a bifactor framework. *Journal of Psychopathology and Behavioral Assessment, 39*, 92–102. doi:10.1007/s10862-016-9565-7.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*, 223–237. doi:10.1080/00223891.2015.1089249.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods, 21*, 137–150. doi:10.1037/met0000045.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*, 243–248. doi:10.1007/s11336-009-9135-y.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: review of practice and implications. *Human Resource Management Review, 18*, 210–222. doi:10.1016/j.hrmr.2008.03.003.

Smith, R. L., Rose, A. J., & Schwartz-Mette, R. A. (2010). Relational and overt aggression in childhood and adolescence: clarifying mean-level gender differences and associations with peer acceptance. *Social Development, 19*, 243–269. doi:10.1111/j.14679507.2009.00541.x.

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

Tackett, J. L., & Ostrov, J. M. (2010). Measuring relational aggression in middle childhood in a multi-informant multi-method study. *Journal of Psychopathology and Behavioral Assessment, 32*, 490–500. doi: 10.1007/s10862-010-9184-7.

Tackett, J. L., Daoud, S. B., De Bolle, M., Burt, S. A. (2013). Is relational aggression part of the externalizing spectrum? A bifactor model of youth antisocial behavior. *Aggressive Behavior, 39*, 149–159. doi: 10.1002/ab.21466.

U.S. Census Bureau (2010). Community facts. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml?src=bkmk.

Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: a Monte Carlo study. *Structural Equation Modeling, 14*, 435–463. doi:10.1080/10705510701301677.