Measurement of aggressive behavior in early childhood: A critical analysis using five informants

Kristin J. Perry [a],[*], Jamie M. Ostrov [a], Dianna Murray-Close [b],
Sarah J. Blakely-McClure [c], Julia Kiefer [a], Ariana DeJesus-Rodriguez [a],
Abigail Wesolowski [a]

[a] University at Buffalo, State University of New York, Buffalo, NY 14260, USA
[b] University of Vermont, Burlington, VT 05405, USA
[c] Canisius College, Buffalo, NY 14208, USA

A R T I C L E   I N F O

A B S T R A C T

Measurement of aggressive behavior in early childhood is unique given that relational aggression is just developing, physical aggression is still prevalent, and both forms of aggression are relatively overt or direct. The current study had three aims. The first aim was to examine the internal reliability, validity, and correspondence of five different assessments of aggressive behavior in early childhood: parent report, teacher report, observer report, child report, and naturalistic school-based observations. The second aim was to test a one- and two-factor model of early childhood aggression using confirmatory factor analysis. The final aim of the study was to investigate gender differences among different reports of aggression. Observations, teacher report, and observer (research assistant) report were collected in the children's school, and parent report and child report were collected in a lab session at one time point (N = 300; 56% male; M_{age} = 44.86 months, SD = 5.55). Observations were collected using a focal child sampling with continuous recording approach, and previously validated measures were used for the remaining four informants. Results demonstrated that all measures were reliable with the exception of child report of relational aggression, and there was small to strong correspondence among the various informants. In addition, a two-factor structure of aggression provided the best fit to the data, providing evidence for divergence among relational and physical aggression. Finally, there were robust gender

 * Corresponding author.
    E-mail address: kperry5@buffalo.edu (K.J. Perry).

differences in physical aggression, but gender differences in relational aggression varied by method. The implications of different types of measurement are discussed.

Published by Elsevier Inc.

Introduction

The study of aggression has a rich history, including distinguishing different forms and functions of aggression and identifying developmental changes in aggression (e.g., Hartup, 1974). Aggression, defined as behaviors intended to hurt or harm an individual, can be enacted using distinct means, including via physical harm (e.g., verbal threats of harm, hitting, kicking) and through damage to relationships (e.g., exclusion, gossiping, verbal threats of withdrawing) (Crick & Grotpeter, 1995; Eisner & Malti, 2015). The majority of early research on aggressive behaviors in childhood focused on elementary school children, with less research among early childhood samples (Landy & Peters, 1992). In addition, studies of aggression in early childhood tended to focus on physical forms of aggressive behavior across a variety of assessment methods, including peer nomination, self-report, and teacher ratings (e.g., Behar & Stringfield, 1974; Johnston, DeLuca, Murtaugh, & Diener, 1977) and, perhaps most commonly, observations (Johnston et al., 1977). Use of these various methods often leads to divergent conclusions about the independence among subtypes of aggressive behavior, gender differences in aggressive behavior, and the various contexts in which aggressive behavior occurs (Card, Stucky, Sawalani, & Little, 2008). The current study evaluated the internal reliability and validity of five different measures of relationally and physically aggressive behavior in early childhood. The ability to distinguish among different informants of aggressive behavior, draw conclusions about the methodological rigor of these various methods, and evaluate the structure of aggressive behavior using multiple methods is critical to accurately conceptualizing aggression in early childhood.

Physical and relational aggression in early childhood

In the research literature, there has been a focus on physical and relational forms of aggression in part because of the historical interest in between- and within-group gender differences in aggression and gender differences in longitudinal studies linking aggression and social-psychological adjustment outcomes (e.g., Crick & Grotpeter, 1995; Crick et al., 2006; Lagerspetz, Björkqvist, & Peltonen, 1988; Murray-Close, Han, Cicchetti, Crick, & Rogosch, 2008). In fact, in early research on the development of aggression in early childhood, the focus on physical and verbal forms of aggression led researchers to conclude that boys were more likely than girls to exhibit aggressive behaviors (e.g., Behar & Stringfield, 1974; Hartup, 1974). However, this conclusion was challenged by research conducted during the 1980s and 1990s, when the study of nonphysical forms of aggression, such as indirect aggression (Lagerspetz et al., 1988), social aggression (Cairns, Cairns, Neckerman, Ferguson, & Gariépy, 1989; Galen & Underwood, 1997), and relational aggression (Crick & Grotpeter, 1995) gained prominence.

The developmental period of early childhood (3–5 years) serves as a critical period for studying the development of physical and relational aggressive subtypes for several reasons. First, forms of aggression may have different correlates and prevalence rates across development (Sijtsema & Ojanen, 2018), highlighting the need for research focused on distinct developmental periods, including early childhood. Second, although both physical and relational aggression are commonly expressed during this period, key developmental changes occur among typically developing children. For instance, children exhibit an increase in physical aggression after the onset of expressive language and then rapid declines as they transition to kindergarten (National Institute of Child Health, Human Development Early Child Care, & Research, 2004; Tremblay et al., 2005). As this developmental shift occurs, relational aggression appears to increase (see Fite & Pederson, 2018).

Third, relational aggression appears to be qualitatively different in early childhood compared to other periods of development (see Casas & Bower, 2018). Specifically, relational aggression in early childhood tends to be direct, the identity of the perpetrator is known, and such aggression is based on the "here and now" and typically does not reflect retaliation for a prior hostile exchange that took place days or weeks earlier (Casas & Bower, 2018). For these reasons, the focus of the current study was on physical and relational aggression during the developmentally salient early childhood period. Given previous research and conceptualizations of physical and relational aggression as related but distinct, it was hypothesized that a two-factor (i.e., physical and relational) model of aggressive behavior with a moderate correlation between the two factors would fit the data better than a one-factor aggression model.

Methods for studying aggression

A critical question for researchers investigating forms of aggression in early childhood is what methods are best suited to capturing these behaviors and whether results differ when different methods are adopted. The most common measures of physical and relational aggression in early childhood include observations, self-report, and other report (e.g., teacher, peer, parent, and observer reports). Importantly, children may behave differently within the various settings they encounter; therefore, assessing behavior in a single setting (e.g., school) might not capture a child's overall aggressive tendencies.

Observations

One of the most rigorous methods for studying aggression is using behavioral observations. Typically, observations in early childhood are conducted within schools because preschool-age children most frequently interact with peers in this setting. Children have low reactivity to observers, particularly when observers are in the classroom for a period of time before the observations (i.e., reactivity period; Ostrov & Keating, 2004) and are trained to be minimally reactive to the children. In the current study, we used a focal child sampling with continuous recording observation procedure (Juliano, Werner, & Cassidy, 2006; Ostrov & Keating, 2004), in which a trained observer conducts an observation on a selected focal child and notes each interaction that the focal child has with other children. Observations are recorded over a set period of time, and the trained observer records interactions within a short distance of the child but does not interact with the child (Juliano et al., 2006; Ostrov & Keating, 2004). Close proximity between observers and focal children allows for the trained observer to differentiate between physical aggression and rough-and-tumble play (Pellegrini, 1989). Observations using the focal child sampling approach have been shown to have good reliability (e.g., Ostrov, Ries, Stauffacher, Godleski, & Mullins, 2008), providing support for the utility of behavioral observations of aggression in early childhood.

Self-report

Child reports are designed to collect children's unique perspective about their own aggressive behaviors across multiple contexts (Godleski & Ostrov, 2020; Phares, Compas, & Howell, 1989) and may be administered within an interview format for young children (e.g., Godleski & Ostrov, 2020). However, there has been some debate about whether young children are able to reliably and validly report on their own behavior (McLeod, Southam-Gerow, & Kendall, 2017). Specifically, they might not have insight into their own behavior, may respond with extremes, or might not report on behavior that is socially undesirable because they want to please adults (Chambers & Craig, 1998; Chambers & Johnston, 2002). Notably, not much work has used child report in early childhood. Previous work examining concordance rates during later developmental periods has found that there are often disagreements among informants. For example, reports from one study found that cross-rater correlations ranged from .10 to .65, with the largest disagreements being between peer assessments and self-reports (Pakaslahti & Keltikangas-Jarvinen, 2000).

Given the limited use of child self-report, particularly in early childhood, further investigation into this methodology is important to evaluate the utility of young children's perspective on their aggressive behavior. Prior work in early childhood using the same child report method used in the current

study has found support for the reliability and validity of this method (Godleski & Ostrov, 2020). Specifically, when using a developmentally appropriate child interview, there was acceptable reliability for relational and physical aggression (Cronbach's α > .68) and evidence of validity, such that child reports of relational and physical aggression were correlated (Godleski & Ostrov, 2020).

Other reporters

Given the challenges of relying on young children as reporters of aggressive behavior and the time-intensive nature of observations, many researchers have used other reporters as informants of a child's aggressive behavior. These include parent, teacher, peer, and observer informants.

Parent. Parent report has historically been a commonly used index of children's behavior problems, including aggressive behavior (Achenbach & Edelbrock, 1978). Parent reports of their children's relational and physical aggression, rated on a Likert-type scale (Casas et al., 2006; Ostrov & Bishop, 2008), may provide details on children's aggressive behavior in a novel context that teachers and observers are not privy to (e.g., with peers on play dates). In addition, parents may be able to recognize more private displays of aggressive behavior, which are harder for teachers and observers to notice (Ostrov & Bishop, 2008; Pellegrini & Bartini, 2000). However, parents might not be as objective as teachers and other observers when reporting on their own children's behavior with peers. In previous studies, parent measures have shown acceptable internal consistency and, in some work, moderate correlations with teacher reports of physical and relational aggression (Ostrov & Bishop, 2008). However, in other studies, parent report was not significantly associated with teacher or observer report, potentially indicating that parents are reporting on aggression outside of the classroom setting (Ostrov, Gentile, & Mullins, 2013).

Teacher. Teacher reports are used to examine the frequency of aggressive behavior displayed within the classroom. Many researchers have used teacher reports of children's aggression (e.g., Johnson & Foster, 2005; McEvoy, Estrem, Rodriguez, & Olsen, 2003; McNeilly-Choque, Hart, Robinson, Nelson, & Olsen, 1996). Previous studies have found teachers to be reliable reporters of relational aggression (Estrem, 2005; Juliano et al., 2006), with weak to strong overlap with observer reports, observations, and peer reports (Crick, Casas, & Mosher, 1997, Crick et al., 2006; Johnson & Foster, 2005). Teachers are also reliable reporters of physical aggression (Crick et al., 1997; Hawley, 2003), with weak to strong correlations with observer reports, behavioral observations, and peer reports (Casas et al., 2006; Estrem, 2005; Ostrov et al., 2008).

One limitation to the use of teacher reports is that they reflect children's behavior in the school environment. In addition, teachers do not always witness the behaviors that occur during peer-to-peer interactions. Despite these limitations, teacher reports of aggressive behavior are one of the most commonly used methods for assessing young children's aggression. Teachers witness more daily peer-to-peer interactions than parents, which may make them better informants of peer-directed aggression than parents. In addition, teachers are generally experienced with children's peer interactions and therefore have an idea of what "typical" behavior is for preschool children. Finally, teachers have consistently been reliable and valid informants of children's aggressive behavior (Estrem, 2005; Johnson & Foster, 2005; Juliano et al., 2006), providing support for the utility of this method.

Other reporter. Other reporters have also been used to examine aggressive behavior in children such as camp counselors and observers (Murray-Close et al., 2008; Ostrov, 2008). For instance, after completing the observations in a classroom where observers are around children for an extended period of time (e.g., 2 or 3 months; Ostrov & Keating, 2004), observers may be randomly assigned to complete "teacher report" measures for students within that classroom (Ostrov, 2008). This method has been reliable in several studies for both physical and relational aggression and has shown small to strong correlations with teacher reports and naturalistic school-based observations of aggressive behavior (e.g., Murray-Close & Ostrov, 2009; Ostrov, 2008). Observer reports of aggressive behavior allow for an objective reporter trained to efficiently identify specific acts of aggression to assess children's aggressive behavior in a classroom context.

Correspondence of measures of aggression

A large extant literature focuses on identifying informant discrepancies, why they exist, and the meaning of such discrepancies (e.g., Achenbach, McConaughy, & Howell, 1987; De Los Reyes, 2013). Recent meta-analytic work has demonstrated that informant agreement is generally larger for more observable behavior when informants are reporting within the same context (e.g., mothers and fathers) and when using continuous measurements (see De Los Reyes et al., 2015). One theoretical model that addresses the interpretation of informant discrepancies is the operations triad model (OTM; De Los Reyes, 2013). The OTM is composed of three parts: *converging operations,* where stronger confidence in results is drawn when there are similarities in relations between target constructs and outcomes, *diverging operations,* where differences in outcomes based on informants are theorized to be meaningful, and *compensating operations,* where differences in informants are due to measurement errors or structural measurement differences (De Los Reyes, 2013). The current study is a precursor to an OTM study with a focus on similarities and differences across informants rather than associations with outcomes.

Specifically, the first aim of the study was to examine the internal reliability and validity of five different methods of aggressive behavior in early childhood: child report, naturalistic observations, observer (research assistant) report, parent report, and teacher report. Based on previous research, it was expected that cross-informant agreement (e.g., converging operations) would be stronger for physical aggression compared with relational aggression because physical aggression is more overt, and therefore more observable, among informants. In addition, it was expected that there would be more agreement among informants in the school context (i.e., teacher report, observations, and observer report) compared with the home context (i.e., parent report) and self-report. Consistent with compensating operations, the reliability was examined for each measure of aggressive behavior and was expected to vary by informant. There was expected to be more convergence for informants using the same measure (i.e., teacher report and observer report) compared with informants using different measures or assessments (i.e., parents vs. teachers, parents vs. observations). Additionally, the second aim of the current study was to evaluate whether a one or two-factor model would fit the data better. Given previous research that has found that the two forms of aggression are related but distinct, it was hypothesized that a two-factor model of aggressive behavior would fit the data best.

The third aim of the study was to evaluate the role of gender in assessments of aggression. In contrast to the common conception of girls as nonaggressive, meta-analytic findings and cross-cultural research suggest that boys and girls engage in similar levels of relational aggression (Card et al., 2008; Lansford et al., 2012). However, gender effects vary by informant; for instance, studies using parent and teacher reports are more likely to report that girls exhibit more relational aggression than boys (Card et al., 2008). Consistent with the extant literature (Card et al., 2008), we hypothesized that between-group analyses would demonstrate higher physical aggression scores among boys than among girls across all measures of aggressive behavior. Prior to evaluating mean gender differences in latent variables of physical and relational aggression, the measurement invariance of the model was tested to evaluate whether the assessments functioned differently for boys and girls in predicting latent factors. Prior research has found mixed evidence for the measurement invariance of different measures across gender, with items functioning differently for boys and girls when using general measures of aggressive behavior (Kim et al., 2010) but with items functioning similarly when evaluating a bifactor model of aggressive behavior (Perry and Ostrov, 2018). Therefore, it was hypothesized that a two-factor confirmatory factor analysis (CFA) of physical and relational aggression would demonstrate strong factorial invariance across gender. In addition, we expected no gender difference in the relational aggression factor but potential gender differences on individual measures of relational aggression such as parent and teacher reports. We further expected that within-group analyses would demonstrate higher relational aggression scores than physical aggression scores among girls across all measures of aggressive behavior. For boys, it was hypothesized that there would be no difference in physical and relational aggression scores or that physical aggression scores would be higher.

The current study

The current research drew on a large multicohort study (N = 300) to examine the reliability and inter-method correspondence for five different methods of assessing aggressive behavior in early childhood: child report, naturalistic observations, observer report, parent report, and teacher report. Classroom-level nesting of different measures of aggressive behavior was also examined. The second goal was to evaluate the structure of aggressive behavior using these different methods. A CFA was conducted to test the factor structure of aggressive behavior (i.e., physical and relational aggression). It was hypothesized that a two-factor structure, with a moderate correlation between the two factors, would fit the data better than a one-factor model. The third aim of the study was to examine the role of gender in aggression by considering the measurement invariance of the final model across gender, between- and within-group gender differences in different measures of aggression, and between-group gender differences in the latent factors of aggression. We hypothesized that there would be within- and between-group gender differences in physical and relational aggression.

Method

Participants

A total of 300 children (44.0% girls; M_{age} = 44.86 months, SD = 5.55) from four cohorts participated in the current study, which is part of a larger study. The sample was somewhat diverse (3.0% African American/Black, 7.6% Asian/Asian American/Pacific Islander, 1.0% Hispanic/Latinx, 11.3% multiracial, 62.1% White, and 15.0% missing/unknown). Parental occupation was gathered at enrollment and was coded using Hollingshead's (1975) four-factor index 9-point scoring system (e.g., 9 = executives and professionals, 1 = service workers). Parents had the opportunity to enter two occupations, 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 7.72 average, indicating that a typical family in our sample was from the second to third highest occupation groups (i.e., 7 = small business owners, farm owners, managers, and minor professionals; 8 = administrators, lesser professionals, and proprietors of medium-sized businesses), which suggests that our sample is on average middle to upper-middle class. Children were recruited from 10 NAEYC (National Association for the Education of Young Children) accredited or recently accredited early childhood education centers. Four of the schools were university affiliated, and six were community based. Education centers participated during different cohorts, providing centers with an opportunity for a break in order to maintain research enthusiasm and minimize teacher and parent burnout throughout the project. Specifically, two schools participated in data collection for all four cohorts, six schools participated for three cohorts, two schools participated for two cohorts, and one school participated for one cohort.

There were complete data for observations and teacher report of aggressive behavior, whereas parent report was available for 155 individuals, observer report was available for 174 individuals, and child report interviews were available for 94 individuals. Analyses of variance (ANOVAs) were run to examine whether missing data were related to any demographic variable (i.e., gender, age, ethnicity, school socioeconomic status [SES], or cohort), observation, or teacher report of aggression. Results demonstrated that children who had missing observer reports were older, $F(1, 294)$ = 7.07, p = .01, R^2 = .02, and had higher physical aggression observation scores, $F(1, 292)$ = 11.77, p = .001, R^2 = .04, and relational aggression observation scores, $F(1, 292)$ = 11.31, p = .001, R^2 = .03. Children who had missing child report data were younger, $F(1, 294)$ = 4.39, p = .04, R^2 = .01, and had higher scores on teacher reports of physical aggression, $F(1, 292)$ = 4.85, p = .03, R^2 = .01. Children who had missing parent report data did not differ on any aggression or demographic variable from children who had parent report data. In addition, due to study design, the cohort that children were in was related to whether data were missing. Therefore, the data were likely missing at random (MAR), and age and cohort were included in the final structural model to help facilitate the full information maximum likelihood (FIML) process (Little, 2013).

Procedures

Data collection occurred when children were 3 years old in the spring and summer of one academic year. Teacher and observer reports occurred toward the end of the spring, when teachers had been the children's teachers for approximately 8 to 9 months (*M* = 8.32 months, *SD* = 2.89, based on 157 children where length of the teacher–child relationship was available) and observers spent 2 to 3 months within the classroom. Parents and children who were involved in the spring data collection were then invited to participate in a summer lab session, where parent and child reports were collected. Parents were also given the option to complete the parent report in the early fall if they were unable to participate in the summer. For participation in the school-based portion of the study, a consent form was distributed to parents in the spring. Parents were compensated $30 to $40 for their time in the lab or for completing a parent report, and children received a small educational toy. Parental consent and child verbal assent were obtained for the lab session. Teacher consent was obtained prior to teacher report completion. Teachers received $5 to $30 based on the number of enrolled children in their classrooms. All procedures in the study were approved by the local institutional review board.

Measures

Early childhood observation system

Trained undergraduate and graduate research assistants collected naturalistic observations using a focal child sampling with a continuous recording procedure (Ostrov & Keating, 2004). Prior to classroom entry, consistent with prior procedures (see Ostrov & Keating, 2004), observers underwent extensive training to identify physical and relational aggression. Typically, there were two or three observers per classroom. Observations were undertaken during a 2-month period, with the goal of completing eight 10-min observation sessions per child. On average, each child had a total of 7.75 (*SD* = 0.77) 10-min observation sessions. Children were included only if they had at least four observation sessions. The sum of physical and relational aggression observations was divided by the number of sessions to get an average aggression score per session to control for any differences in the number of sessions children had. Reliability sessions were collected for 16.5% of all sessions spread across collection, which is within an acceptable range of inter-rater reliability sampling percentages (e.g., Ostrov & Hart, 2013). These sessions demonstrated that observations of relational aggression (intraclass correlation coefficient [ICC] = .78) and physical aggression (ICC = .81) were reliable. In addition, reactivity (i.e., the frequency with which focal children commented, looked, or asked questions to observers) per observation session was low (*M* = 0.36, *SD* = 0.36) and was comparable to previous studies (Ostrov et al., 2008; Perry & Ostrov, 2018).

Preschool Social Behavior Scale–Teacher Form

The Preschool Social Behavior Scale–Teacher Form (PSBS-TF) was used to assess teacher reports of children's aggression (Crick et al., 1997). This measure includes 6 items that assess physical aggression (e.g., "This child hits or kicks others") and 6 items that assess relational aggression (e.g., "This child tells others not to play with or be a peer's friend"). This measure uses a 5-point Likert scale ranging from 1 (*never or almost never true*) to 5 (*always or almost always true*), and a weighted sum was used to get an index of aggressive behavior. Prior work has supported the reliability of the PSBS-TF (e.g., Crick et al., 1997; Ostrov, 2008; Ostrov & Keating, 2004; Perry & Ostrov, 2018) and overlap with naturalistic observations and observer reports (Crick et al., 2006; Ostrov, 2008; Ostrov & Keating, 2004; Perry & Ostrov, 2018). In addition, factor analyses have supported a delineation among the relational and physical aggression factors (Crick et al., 1997; Perry & Ostrov, 2018). In the current sample, the physical and relational aggression subscales were reliable, with Cronbach's *α*s of .87 and .92, respectively.

Preschool Social Behavior Scale–Observer Report

The Preschool Social Behavior Scale–Observer Report (PSBS-OR; Ostrov, 2008) was used to assess observer reports of children's physical and relational aggression. The observer report version of the PSBS is a parallel version of the PSBS-TF and was used by the same trained research assistants who

conducted previous naturalistic observations of the children. After completing the observations in a classroom, observers were randomly assigned to complete measures for students within that classroom. The PSBS-OR has been shown to have good reliability and moderate to high correlations with teacher reports of the PSBS and naturalistic observations (Murray-Close & Ostrov, 2009; Perry & Ostrov, 2018). In the current study, the measure demonstrated good reliability for relational aggression (Cronbach's α = .90) and physical aggression (Cronbach's α = .86).

Children's Social Behavior–Parent Report

The Children's Social Behavior–Parent Report (CSB-PR) is a 13-item parent report that assesses children's physical aggression (4 items; e.g., "Hits or kicks other kids") and relational aggression (5 items; e.g., "When mad at other kids, gets even by excluding those kids from his/her play group or friendship group"), with responses ranging from 1 (*never true*) to 5 (*almost always true*). Positively toned items are also included in this measure (prosocial behaviors; 4 items) but were not used in the current study. Previous research has found acceptable reliability and validity for both subscales (Ostrov & Bishop, 2008; Ostrov et al., 2013). In this study, for the paired-samples *t* tests, the items were averaged for both subscales of physical and relational aggression so that they would be on the same metric. However, for the CFA, the sums of the items were used. This study found acceptable reliability for both relational aggression (Cronbach's α = .72) and physical aggression (Cronbach's α = .83).

Child Social Behavior Scale–Revised

The Child Social Behavior Scale–Revised (CSBS-R) was used to assess child-reported aggression. This initial measure was developed from the original CSBS reported by Crick and Grotpeter (1995) that was used for children in middle childhood. The CSBS-R is a 12-item interview that asks children about their aggressive and prosocial behaviors (Godleski & Ostrov, 2020). In this study, both the physical aggression subscale (4 items) and relational aggression subscale (4 items) of the measure were used. Each item was asked using a developmentally appropriate response scale within the context of an individual interview with a graduate research assistant who had first established rapport with children. Rapport was built prior to the interview by introductions with children, followed by a non-interview activity such as coloring with children while discussing their interests. Children were presented with a board that had a race track and stoplight on it. They were then instructed what each light meant (red light: "no"; yellow light: "yes, a little"; green light: "yes, a lot"). Children used a toy car or pointed to the color to provide their answer. Children could also provide their answer verbally, and the interviewer always verbally confirmed children's answers. Previous research has found acceptable reliability for the aggression scales (Godleski & Ostrov, 2020). In the current study, the relational aggression subscale was not reliable (Cronbach's α = .53). Given the low reliability, removing the first item of the measure was considered because previous work had demonstrated that this improved reliability (Godleski & Ostrov, 2020). However, further analysis showed that removing any item would reduce the reliability. Therefore, the relational aggression subscale was included in preliminary analyses but was not included in the CFA. For physical aggression, acceptable reliability was found (Cronbach's α = .71).

Data analysis

First, descriptive data and correlations of the measures were obtained. Outliers were modified by adjusting the value to ± 3 standard deviations from the mean, and for key study variables skew values (0.79–2.02) and kurtosis values (0.10–5.33) were within accepted ranges for normally distributed variables (Kline, 2016).

To address the first aim of this study, bivariate correlations were examined to determine the correspondence among different measures of aggressive behavior. ICCs were examined for each measure to determine whether there was variance at the classroom level.

To evaluate the second aim, a CFA was used to test the one- and two-factor models of aggressive behavior. First, a one-factor model was tested where all indicators of aggressive behavior loaded on an

overall aggressive behavior factor. Second, a two-factor model was tested where the relational aggression indicators were specified to load on a relational aggression factor and physical aggression indicators were specified to load on a physical aggression factor with a correlation between the two factors. Error correlations between physical and relational aggression indicators within method were then estimated. It was expected that the first two models would provide a poor fit to the data given that high overlap between relational and physical aggression within each measure was expected. Next, two models were run to check the robustness of the overall model given missing data for parent and observer reports, namely (a) a CFA was conducted using the sample with full parent report (n = 155) to evaluate whether missing data played a role in model fit and (b) a CFA was evaluated using only observer report, observations, and teacher report (n = 174) to determine whether including parent report changed results.

To test the third aim of the study, repeated-measures ANOVAs were used to determine whether differences between relational and physical aggression varied across gender. Between-group gender differences were examined using ANOVAs to determine whether girls or boys had higher aggression scores for each measure. In addition, the measurement invariance of the final CFA was tested across gender. Finally, a structural model was tested where gender was added as a predictor of the final measurement model.

SPSS was used for preliminary analyses, reliability, and bivariate correlations. Mplus Version 8.6 (Muthén & Muthén, 1998) was used to run the CFAs and ICCs as well as the structural model. Maximum likelihood estimation with robust standard errors (MLR) was used due to a slight skew in some of the variables. Missing data were accommodated by using FIML. Demographic covariates were included in the model only if they were related to outcomes at $p < .05$. School SES, age, and cohort were included in the model to accommodate the FIML process and serve as covariates. To test the measurement invariance of the model across gender, three models were run: (a) a configural model where all paths were freed across gender, (b) a metric model where factor loadings were constrained to equivalence across gender, and (c) a scalar model where factor loadings and intercepts were constrained to equivalence across gender. Methods developed by Satorra and Bentler (2010) were used to calculate the χ^2 difference test statistic using the MLR estimator. Children were clustered within classrooms (n = 50), which was controlled for in the measurement and structural models.

A likelihood ratio χ^2 test was used to test overall model fit where $p > .05$ indicated good model fit. The following alternative fit indices were also considered: (a) comparative fit index (CFI), where values greater than .95 suggest good fit, (b) standardized root mean square residual (SRMR), where values less than .08 represent mediocre fit and values less than .05 indicate close fit, and (c) root mean square error of approximation (RMSEA), where values less than .08 suggest mediocre fit and values less than .05 indicate close fit (Hu & Bentler, 1999).

Results

Measure characteristics

Inter-method correspondence

In regard to relational aggression, observations, observer reports, and teacher reports were positively correlated (rs = .21–.40, $p < .01$) (see Table 1). Parent and child reports of relational aggression were not significantly correlated with the other informants but were significantly correlated with each other (r = .31, $p < .001$). In regard to physical aggression, observations, observer reports, and teacher reports were positively correlated (rs = .40–.41, $p < .001$). Parent reports were significantly associated with observer and teacher reports (rs = .43 and .37, respectively, ps < .01). Parent reports were not associated with observations (r = .07), and child reports were not related to any other report of physical aggression. See the online supplementary material for paired-samples t tests that demonstrate whether physical or relational aggression scores were higher for each measure and for information about correspondence among observations and teacher reports for children demonstrating high levels of aggressive behavior.

Table 1
Full sample descriptive statistics and correlations.

	1	2	3	4	5	6	7	8	9	10
1. Ragg TR	–					.54**	.14+	.12*	.16*	− .07
2. Ragg OR	.25**	–				.07	.51**	.16*	.27**	− .08
3. Ragg Obs	.21**	.40**	–			.10+	.16*	.27**	− .07	− .15
4. Ragg PR	.16+	.09	.06	–		.09	−.02	−.08	.37**	.10
5. Ragg CR	−.03	.05	−.12	.31**	–	.01	−.08	−.16	.18+	.45**
6. Pagg TR						–				
7. Pagg OR						.41**	–			
8. Pagg Obs						.41**	.40**	–		
9. Pagg PR						.37**	.43**	.07	–	
10. Pagg CR						.17	−.13	−.09	.13	–
M	9.82	8.29	0.13	8.48	1.67	8.85	8.11	0.26	5.63	0.71
SD	4.61	3.43	0.17	2.95	1.84	3.48	2.98	0.31	2.14	1.45
Range	6.0–23.88	6.0–19.95	0.0–0.70	5.0–17.50	0.0–7.00	6.0–20.04	6.0–17.70	0.0–1.83	4.0–12.08	0.0–5.40

Note. Ragg, relational aggression; TR, teacher report; OR, observer report; Obs, observations; PR, parent report; CR, child report; Pagg, physical aggression. Inter-rater correspondence within each subtype of aggression is shown below the diagonal; correlations between relational and physical aggression are shown above the diagonal, with the correlation between physical and relational aggression by method bolded.

+ $p < .10$.
* $p < .05$.
** $p < .01$.

Intraclass correlation coefficients

ICCs were used to examine the amount of variability in each measure that was attributable to the classroom level (see Table 2). ICCs for measures of relational aggression ranged from .04 to .33, and ICCs for measures of physical aggression ranged from .00 to .22.

Confirmatory factor analysis

A CFA was conducted examining a one-factor model of aggressive behavior, where parent report, teacher report, observer report, and observations were used as indicators, controlling for classroom membership. Child report was not included due to low sample size and poor correspondence with other methods. The model provided a poor fit to the data, $\chi^2(20) = 389.62$, $p < .001$, CFI = .00, RMSEA = .25, SRMR = .11. Next, a two-factor model was estimated, where the four physical aggression indicators loaded on one factor and the four relational aggression indicators loaded on a second factor with a correlation between the two factors, controlling for classroom membership. The model pro-

Table 2
Intraclass correlation coefficients by classroom.

	Intraclass correlation coefficient
Relational aggression	
Teacher ratings	.33
Observer ratings	.29
Parent ratings	.04
Observations	.19
Physical aggression	
Teacher ratings	.22
Observer ratings	.12
Parent ratings	.00
Observations	.19

Note. Child report was not included because of a low number of individuals. The number of clusters ranged from 33 to 50, and the average number of children per cluster ranged from 3.23 to 5.88.

vided a poor fit to the data, $\chi^2(19) = 294.06$, $p < .001$, CFI = .01, RMSEA = .22, SRMR = .10, but was a better fit to the data than the one-factor model, $\Delta\chi^2(1) = 95.56$, $p = .001$. Factor loadings ranged from .18 to .81 on the relational aggression factor and from .42 to .86 on the physical aggression factor, with a strong correlation between the two factors ($r = .70$, $p < .001$). Next, the error correlations between indicators from the same method were allowed to correlate to account for shared method variance. This model provided a good fit to the data, $\chi^2(15) = 17.67$, $p = .28$, CFI = .99, RMSEA = .02, SRMR = .05 (see Fig. 1). Factor loadings ranged from .18 to .75 on the relational aggression factor and from .49 to .68 on the physical aggression factor, with a moderate correlation between the two factors ($r = .35$, $p = .03$), suggesting that once accounting for shared method variance, the correlation between physical and relational aggression is attenuated. Parent report of relational aggression did not significantly load on the relational aggression factor, but all other loadings were significant. There were significant correlations between the errors for each method, with the smallest correlation for observations ($r = .24$, $p < .001$) indicating that there is the least amount of shared method variance when using observations. There were higher error correlations for parent report ($r = .40$, $p < .001$), teacher report ($r = .64$, $p < .001$), and observer report ($r = .65$, $p < .001$).

Two models that tested the robustness of the overall model were examined. The two-factor model of aggression with correlated errors was rerun in the subsample with full parent report ($n = 155$) and provided an acceptable fit to the data, $\chi^2(15) = 21.10$, $p = .13$, CFI = .97, RMSEA = .05, SRMR = .06. Factor loadings were similar, with relational aggression loadings ranging from .17 to .94 and physical aggression loadings ranging from .47 to .78. Next, the same model was run with full observer report and without the parent report indicators ($n = 174$). This model provided a good fit to the data, $\chi^2(5) = 1.83$, $p = .87$, CFI = 1.00, RMSEA = .00, SRMR = .03. Loadings were similar, with observer report having the highest loading on both factors.

Gender differences

Measurement invariance

First, the measurement invariance of the model across gender was tested. The configural invariance model, $\chi^2(30) = 46.20$, $p = .03$, CFI = .95, RMSEA = .06, SRMR = .08, and the metric invariance model, $\chi^2(36) = 44.69$, $p = .15$, CFI = .97, RMSEA = .04, SRMR = .08, provided an acceptable fit to the data, with no difference in model fit between the two models, $\Delta\chi^2(6) = 2.12$, $p = .91$. A scalar invariance model provided an acceptable fit to the data, $\chi^2(42) = 52.55$, $p = .13$, CFI = .97, RMSEA = .04, SRMR = .09, and
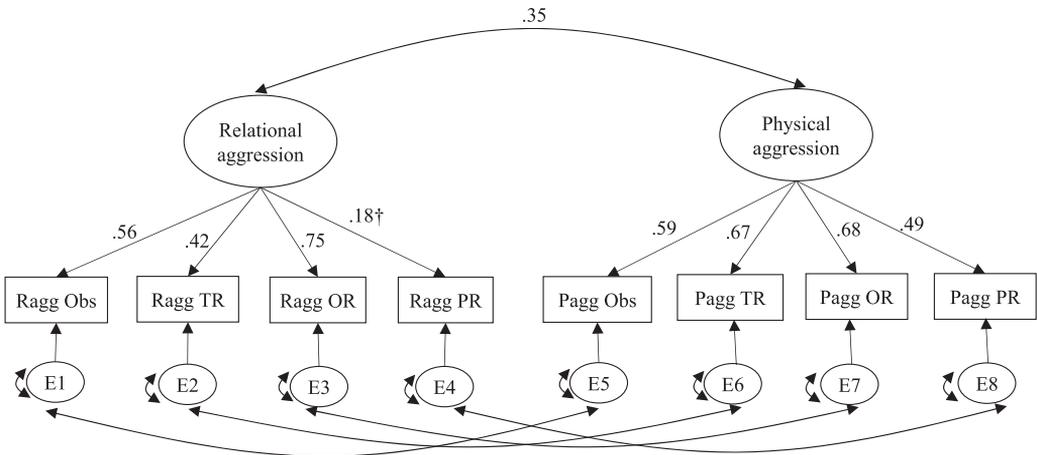

Fig. 1. Two-factor model of aggressive behavior. This model shows the final best-fitting two-factor model of aggressive behavior in the full sample. A dagger (†) indicates that the loading was not significant; all other loadings were significant at $p < .05$. Ragg, relational aggression; Pagg, physical aggression; TR, teacher report; PR, parent report; Obs, observations; OR, observer report; E, Error term. Child report was not included in the model because of poor reliability and poor correspondence with other raters.

provided no difference in model fit with the metric invariance model, $\Delta\chi^2(6) = 7.84$, $p = .25$. Therefore, the two-factor model of aggressive behavior demonstrated strong factorial invariance across gender.

Differences in measures across gender

Repeated-measures ANOVAs were used to determine whether the magnitude of within-group gender differences was significant across genders. A main effect of aggression (relational vs. physical), a main effect of gender, and their interaction (which is of interest in the current analyses) were tested. For child report, both girls and boys were significantly more likely to endorse relational aggression relative to physical aggression, and this effect did not differ by gender, $\Lambda = .99$, $F(1, 91) = 0.68$, $p = .41$. For parent report, $\Lambda = .95$, $F(1, 152) = 8.61$, $p = .004$, teacher report, $\Lambda = .85$, $F(1, 292) = 52.90$, $p < .001$, observer report, $\Lambda = .90$, $F(1, 172) = 20.19$, $p < .001$, and observations, $\Lambda = .96$, $F(1, 292) = 12.01$, $p = .001$, there was a significant type of aggression by gender interaction. Specifically, parents reported that relational aggression was significantly more prevalent than physical aggression, particularly among girls. Teachers rated girls as higher in relational aggression compared with physical aggression and reported that boys engaged in similar levels of physical and relational aggression. Observers rated girls as exhibiting higher relational aggression scores compared with physical aggression scores, whereas they rated boys as having significantly higher physical aggression scores compared with relational aggression scores. Finally, physical aggression observations were higher than relational aggression observations for both genders, but this difference was larger for boys relative to girls.

Between-group gender differences

Univariate ANOVAs were used to examine between-group gender differences in aggressive behavior (Table 3). In regard to physical aggression, boys had significantly higher scores than girls when using teacher report, observer report, parent report, or observations (R^2 values ranged from .01 to .07). There was no difference in physical aggression when using child report. In regard to relational aggression, girls had significantly higher scores than boys when using teacher report or observations. There was no difference in scores when using child report, parent report, or observer report.

Structural model

A path analysis was tested where the latent factors were regressed on gender and the covariates (i.e., school SES, age, and cohort), controlling for classroom membership. The robust χ^2 could not be estimated with the dummy-coded cohort variables in the model, so these were dropped. This model provided an acceptable fit to the data, $\chi^2(33) = 75.35$, $p < .001$, CFI = .89, RMSEA = .07, SRMR = .075. Gender (1 = boys, 2 = girls) was associated with physical aggression ($\beta = -.35$, $p < .001$), such that boys

Table 3
Between-group gender differences in aggression by rater.

Informant	F	df	p	Adjusted R^2
Relational aggression				
Teacher ratings	8.83	1, 291	.003	.03
Observer ratings	1.35	1, 172	.45	.002
Parent ratings	0.02	1, 151	.88	.00
Child report	0.38	1, 91	.54	.00
Observations	3.81	1, 291	.05	.01
Physical aggression				
Teacher ratings	15.71	1, 291	<.001	.05
Observer ratings	11.20	1, 172	.001	.06
Parent ratings	12.37	1, 151	.001	.07
Child report	0.04	1, 91	.83	.00
Observations	4.56	1, 291	.03	.01

Note. $p < .05$ indicates a significant between-group gender difference. Univariate tests were used to test unique gender effects. See text for interpretation of gender differences.

had higher physical aggression scores than girls, but it was not associated with relational aggression (β = .18, p = .12).

Discussion

There were three aims of the current study. The first aim was to examine the internal measurement characteristics of different measures of aggressive behavior. Results demonstrated that all methods were reliable with the exception of child report of relational aggression. In addition, as expected, inter-rater correspondence of aggression was weak to strong, with greater correspondence when using physical aggression and the weakest correspondence between parents and other raters. The second aim of the study was to evaluate the structure of aggressive behavior using these different methods. A CFA demonstrated that a two-factor structure of physical and relational aggression best represented the data, with error correlations accounting for shared method variance consistent with hypotheses. The third aim was to examine the role of gender in aggression by considering the measurement invariance of the final model across gender, between- and within-group gender differences in different measures of aggression, and between-group gender differences in the latent factors of aggression. In the final structural model, boys had higher physical aggression scores than girls, but there was no difference in relational aggression scores, consistent with hypotheses. There were robust within-group differences for girls, such that they generally demonstrated higher relational aggression scores relative to physical aggression scores across method.

In general, the measurement characteristics of the various methods were favorable with the exception of child report of relational aggression, although it has been reliable in past studies (i.e., Godleski & Ostrov, 2020). The development of executive functioning, self-development, or theory of mind may play an important role in a child's ability to report on more complex behaviors such as relational aggression (Karabenick et al., 2007; Woolley, Bowen, & Bowen, 2016). Despite the lack of reliability, child self-report may be particularly important when trying to understand the function of a child's aggressive behavior. How a young child thinks about aggressive behavior, regardless of accuracy of frequency reports, may be important to consider for intervention. Further work should investigate why young children might not report their aggression and how measurement approaches can mitigate social desirability (Pakaslahti & Keltikangas-Jarvinen, 2000). Efforts to build rapport with children by engaging them in their natural environments prior to assessing their behavior may mitigate social desirability and lead to increased reliability and validity.

Inter-rater correspondence of aggression across the measures was weak to strong, with greater correspondence when using physical aggression and the weakest correspondence between parents and other raters. The greater correspondence for physical aggression compared with relational aggression is consistent with prior meta-analytic work demonstrating that more observable behavior (i.e., externalizing vs. internalizing behavior) has greater informant correspondence (De Los Reyes et al., 2015). In addition, given safety concerns, physical aggression incidents at school may be communicated more frequently to parents relative to relational aggression. Furthermore, teachers and parents have a lower level of tolerance and are more likely to intervene when physical aggression is being displayed relative to relational aggression (Coplan, Bullock, Archbell, & Bosacki, 2014; Swit, McMaugh, & Warburton, 2018; Werner, Senich, & Przepyszny, 2006). Findings underscore potential strengths and limitations of approaches to measuring both physical and relational aggression in early childhood and suggest that measurement is likely to be strengthened by the inclusion of multiple informants and approaches.

The weaker correspondence between parents and other raters may have occurred for two reasons. First, as hypothesized in the diverging operations portion of the OTM (De Los Reyes, 2013), the absence of associations across informants may in part reflect the context in which aggressive behavior is displayed (e.g., school-based peers, neighborhood friends, siblings). Indeed, each informant may be privy to different displays of aggressive behavior. Public and private assessments often provide unique insights regarding children's behavior, and objective behavioral observations might not converge with child interviews or parent reports for this reason (Pellegrini & Bartini, 2000). In addition, given the shared context, school-based behavioral observations of aggression are likely to be more strongly related to teacher reports than to parent reports (De Los Reyes, 2013; Ostrov & Bishop, 2008). In fact,

prior research has found that when using the same observational paradigm in different contexts, pre-school children display context-specific disruptive behavior, suggesting that there are likely real differences in behavior in the home and school contexts (De Los Reyes, Henry, Tolan, & Wakschlag, 2009). Second, consistent with the compensating operations facet of the OTM (De Los Reyes, 2013), which encompasses measurement error and differences in measures, it was expected that parents would have lower convergence because they had a different aggression measure than teachers or observers. Therefore, contextual differences, measurement differences, and other sources of error all are likely contributing to informant differences.

Finally, results from the clustering analysis (i.e., ICCs) demonstrated that there was classroom clustering for all raters except for parent report of physical aggression. Classroom effects were larger for relational aggression than for physical aggression for parent, teacher, and observer reports, whereas classroom effects did not differ for observations of physical and relational aggression. Therefore, even when a different rater is used for each participant (i.e., parent report), there is still some meaningful variability at the classroom level for relational aggression, perhaps highlighting how classroom norms and behaviors may foster engagement in relational aggression relative to physical aggression. These results suggest that it would be beneficial to control for classroom variability when possible.

The second aim of the study was to examine a CFA of aggressive behavior. Results indicated that a two-factor structure of aggressive behavior best represented aggression in our early childhood sample. This is consistent with prior factor analyses finding that aggression items load on two factors (e.g., Crick et al., 1997; Perry & Ostrov, 2018). This study expands that prior work by confirming a two-factor structure using multiple informants of aggressive behavior. All assessments significantly loaded on their respective factors with the exception of parent report of relational aggression, demonstrating that parents may be reporting on aggression in situations outside of the school environment and might not be the best reporters of peer-based relational aggression (Pellegrini & Bartini, 2000). In addition, results demonstrated that once accounting for shared method variance, the correlation between physical aggression and relational aggression was moderate, indicating that shared method variance may be one of the main driving factors in high correlations between the subtypes of aggression.

The final aim of the current study was to evaluate between- and within-group gender differences in aggression. In the final structural model, boys had higher physical aggression scores than girls, but there was no gender difference in relational aggression. When evaluating gender differences within each method, gender differences were fairly robust for physical aggression, which is consistent with meta-analytic work (Card et al., 2008). Gender differences were more variable for relational aggression; specifically, girls had higher scores when using observations or teacher report, but not when using child, parent, or observer report, suggesting that methodology and informant may be driving gender differences of relational aggression. This is consistent with reviews and meta-analyses of the literature that have found null or mixed results for between-group gender differences in relational aggression with the exception of certain informants such as teachers (Archer & Coyne, 2005; Card et al., 2008). More work is needed to determine why gender differences in relational aggression occur when certain raters are used.

In addition, within-group gender differences are often ignored in discussions of gender differences in aggressive behavior. Prior work has found that girls tend to use relational aggression (and other nonphysical forms of aggression) more than physical aggression (see Archer & Coyne, 2005). Evidence to support this assertion was found in the current study. Specifically, for girls, relational aggression was rated as more frequent than physical aggression for all raters, including observers. For observations, physical aggression scores were higher than relational aggression scores for both genders, but this difference was smaller for girls than for boys. Overall, results from these analyses underscore the importance of considering gender in developmental models of aggressive behavior.

Limitations

Despite a number of strengths to the current study, such as a large sample size and a comprehensive examination of the measurement characteristics of multiple methods, there are also several limitations. First, although five methods were used to assess aggression, additional assessment techniques are available that were not included in the current study. For instance, alternative obser-

vational paradigms, such as the scan sampling approach (Ladd, Price, & Hart, 1988; McNeilly-Choque et al., 1996) may also be useful for observing aggressive behavior. In addition, there was no assessment of peer report of aggressive behavior. Peer nominations and ratings are used to examine the perception of children's aggressive behavior by their peers. Researchers have generally found that young children provide reliable peer reports of overt/physical aggression and that these reports are generally moderately correlated with teacher reports and observations (Crick et al., 2006; Johnson & Foster, 2005) but that peers are less reliable informants of relational aggression (Crick et al., 1997; McNeilly-Choque et al., 1996). Nonetheless, peer reports may offer a unique perspective on children's aggressive behavior when developmentally appropriate assessments are used.

Second, because the sample was focused on early childhood, it is not clear whether similar patterns of findings would emerge during other developmental periods. Similarly, participants were from a limited number of schools from one area in the northeastern United States. Prior work has found that there are mean levels of difference in aggression across cultures (Lansford et al., 2012), and therefore the results might not be generalizable to other geographic regions or cultures.

In addition, the current study focused on physical and relational aggression and did not include other subtypes of aggression such as verbal aggression (e.g., insults, mean names) and nonverbal aggression (i.e., aggressive gestures and nonverbal hostile behaviors but not malicious ignoring, which is a part of relational aggression) (Ostrov & Keating, 2004). It is less common to examine the separate construct of verbal aggression because verbal threats of physical harm are often included in the definition of physical aggression (Crick et al., 2006), verbal threats are common exemplars of relational aggression (e.g., "You can't come to my birthday party"), and boys and girls engage in relatively equal amounts of the behavior during this period of development (Ostrov & Keating, 2004). Nonetheless, future research should explicitly examine additional subtypes of aggression.

Finally, the study focused on evaluating the internal reliability, validity, and structure of aggression using a cross-sectional design. Future work should address the longitudinal measurement characteristics of various measures of aggressive behavior and the broader reliability and validity of latent constructs of aggressive behavior.

Future directions

Overall, results from this study support the utility of multiple methods for studying aggressive behavior in early childhood. Specifically, parent, teacher, and observer reports, as well as observations, all significantly contributed to the physical and relational aggression factors with the exception of parent report of relational aggression, which did not significantly load on the factor. This suggests that having informants from different contexts may provide a more accurate understanding of children's aggression across multiple contexts (i.e., children's aggression with peers at school and with peers in the home). However, it is not always feasible for researchers to use multiple methods. When researchers are able to use only one method, we encourage them to think about the context they are interested in studying and how their chosen informant may affect the interpretation of the results. For example, if researchers are interested in examining peer relations outside of schools, parent report may be most beneficial, whereas if they are interested in evaluating the impact of a classroom intervention where teachers are the interventionists, observations or observer report may be most useful to obtain objective measurements of classroom behavior.

Moreover, consistent with the OTM (De Los Reyes, 2013), future research should examine how cross-informant differences in aggression incrementally predict social-psychological outcomes. The OTM offers a framework for meaningfully studying and understanding informant discrepancies, and there is a variety of statistical methods that allow researchers to parse apart informant differences. For example, the rater trifactor model is a useful way to parse apart measurement differences within an individual while also accounting for differences between raters who have evaluated a set of individuals (see Shin, Rabe-Hesketh, & Wilson, 2019). The next step in aggression methodology research should examine how informant differences contribute to the prediction of outcomes for different forms of aggression. Results from this study suggest that there are measurement differences in physical and relational aggression. For example, gender differences vary by method, raters are more likely to endorse relational aggression compared with physical aggression, and clustering at the classroom

level is generally stronger for relational aggression compared with physical aggression. These results emphasize that the characteristics of the measures or constructs may vary for the different subtypes of aggressive behavior, underscoring the importance of examining both aggressive subtypes. It is also critical that researchers include measurements of the function of aggressive behavior in future work to evaluate how proactive or reactive relational and physical aggression (see Eisner & Malti, 2015) may change informant agreement and discrepancies. Researchers should also strive to be aware of these measurement issues and address them in analyses if possible (e.g., accounting for clustering within classrooms). Future research should examine how measurement differences change over time by examining the measurement invariance of these measures across time as well as how the impact of raters changes as children get older.

In conclusion, a number of different methods demonstrated favorable measurement characteristics. In general, the different informants had moderate to high levels of convergence, with greater convergence among measures of physical aggression. When examining the structure of aggressive behavior, results supported a two-factor model of aggressive behavior, once again demonstrating the uniqueness of physical and relational aggression in early childhood. Notably, the relation between physical and relational aggression was attenuated when accounting for shared method variance. Finally, results suggest that in early childhood there are robust between-group gender differences in physical aggression, but between-group gender differences in relational aggression vary by method. These results expand prior work by testing these differences by rater and using a latent variable approach. There was also evidence of within-group gender differences for physical and relational aggression, underscoring the importance of including gender-informed models of relational aggression. Overall, results from this study demonstrate the utility of multiple methods in the study of aggressive behavior in early childhood, emphasize the importance of considering both physical and relational aggression when studying aggressive behavior, and illustrate the multiple ways in which gender may play a role in the forms of aggressive behavior.

Acknowledgments

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2021.105180.

References

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*, 1275–1301.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213–232.

Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review, 9*, 212–230.

Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601–610.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Ferguson, L. L., & Gariépy, J.-L. (1989). Growth and aggression: I. Childhood to early adolescence. *Developmental Psychology, 25*, 320–330.

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.

Casas, J. F., & Bower, A. A. (2018). Developmental manifestations of relational aggression. In S. M. Coyne & J. M. Ostrov (Eds.), *The development of relational aggression* (pp. 29–48). New York: Oxford University Press.

Casas, J. F., Weigel, S. M., Crick, N. R., Ostrov, J. M., Woods, K. E., Yeh, E. A. J., & Huddleston-Casas, C. A. (2006). Early parenting and children's relational and physical aggression in the preschool and home contexts. *Journal of Applied Developmental Psychology, 27*, 209–227.

Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain, 78*, 27–37.

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36.

Coplan, R. J., Bullock, A., Archbell, K. A., & Bosacki, S. (2014). Preschool teachers' attitudes, beliefs, and emotional reactions to young children's peer group behaviors. *Early Childhood Research Quarterly, 30*, 117–127.

Crick, N. R., Casas, J. F., & Mosher, M. (1997). Relational and overt aggression in preschool. *Developmental Psychology, 33*, 579–588.

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

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

De Los Reyes, A. (2013). Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology, 25*, 669–682.

De Los Reyes, A., Augenstein, T. M., Wang, M. O., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*, 858–900.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology, 37*, 637–652.

Eisner, M. P., & Malti, T. (2015). Aggressive and violent behavior. In M. E. Lamb & R. M. Lerner (Eds.), Handbook of child psychology and developmental science: Socioemotional processes (Vol. 3, 7th ed., pp. 794–841). Hoboken, NJ: John Wiley.

Estrem, T. L. (2005). Relational and physical aggression among preschoolers: The effect of language skills and gender. *Early Education & Development, 16*, 207–232.

Fite, P. J., & Pederson, C. A. (2018). Developmental trajectories of relational aggression. In S. M. Coyne & J. M. Ostrov (Eds.), The development of relational aggression (pp. 49–60). New York: Oxford University Press.

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*, 589–600.

Godleski, S. A., & Ostrov, J. M. (2020). Parental influence on child report of relational attribution biases during early childhood. *Journal of Experimental Child Psychology, 192* 104775.

Hartup, W. W. (1974). Aggression in childhood: Developmental perspectives. *American Psychologist, 29*, 336–341.

Hawley, P. H. (2003). Strategies of control, aggression and morality in preschoolers: An evolutionary perspective. *Journal of Experimental Child Psychology, 85*, 213–235.

Hollingshead, A. B. (1975). Four factor index of social status. Unpublished working paper, Department of Sociology, Yale University.

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.

Johnson, D. R., & Foster, S. L. (2005). The relationship between relational aggression in kindergarten children and friendship stability, mutuality, and peer liking. *Early Education and Development, 16*, 141–160.

Johnston, A., DeLuca, D., Murtaugh, K., & Diener, E. (1977). Validation of a laboratory play measure of child aggression. *Child Development, 48*, 324–327.

Juliano, M., Stetson Werner, R., & Wright Cassidy, K. (2006). Early correlates of preschool aggressive behavior according to type of aggression and measurement. *Journal of Applied Developmental Psychology, 27*, 395–410.

Karabenick, S., Woolley, M., Friedel, J., Ammon, B., Blaxevskil, J., Bonney, C., ... Kelly, K. (2007). Cognitive processing of self report items in educational research: Do they think what we mean? *Educational Psychologist, 42*, 139–151.

Kim, S., Kim, S.-H., & Kamphaus, R. W. (2010). Is aggression the same for boys and girls? Assessing measurement invariance with confirmatory factor analysis and item response theory. *School Psychology Quarterly, 25*, 45–61.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford.

Ladd, G. W., Price, J. M., & Hart, C. H. (1988). Predicting preschoolers' peer status from their playground behaviors. *Child Development, 59*, 986–992.

Lagerspetz, K. M. J., Björkqvist, K., & Peltonen, T. (1988). Is indirect aggression typical of females? Gender differences in aggressiveness in 11- to 12-year-old children. *Aggressive Behavior, 14*, 403–414.

Landy, S., & Peters, R. D. (1992). Toward an understanding of a developmental paradigm for aggressive conduct problems during the preschool years. In R. D. Peters, R. J. McMahon & V. L. Quinsey (Eds.), Aggression and violence throughout the life span (pp. 1–30) Newbury Park, CA: Sage.

Lansford, J. E., Skinner, A. T., Sorbring, E., Giunta, L. D., Deater-Deckard, K., Dodge, K. A., ... Chang, L. (2012). Boys' and girls' relational and physical aggression in nine countries. *Aggressive Behavior, 38*, 298–308.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York: Guilford.

McEvoy, M. A., Estrem, T. L., Rodriguez, M. C., & Olson, M. L. (2003). Assessing relational and physical aggression among preschool children: Intermethod agreement. *Topics in Early Childhood Special Education, 23*(2), 51–61.

McLeod, B. D., Southam-Gerow, M. A., & Kendall, P. C. (2017). Observer, youth, and therapist perspectives on the alliance in cognitive behavioral treatment for youth anxiety. *Psychological Assessment, 29*, 1550–1555.

McNeilly-Choque, M. K., Hart, C. H., Robinson, C. C., Nelson, L. J., & Olsen, S. F. (1996). Overt and relational aggression on the playground: Correspondence among different informants. *Journal of Research in Childhood Education, 11*, 47–67.

Murray-Close, D., Han, G., Cicchetti, D., Crick, N. R., & Rogosch, F. A. (2008). Neuroendocrine regulation and physical and relational aggression: The moderating roles of child maltreatment and gender. *Developmental Psychology, 44*, 1160–1176.

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.

Muthén, L. K., & Muthén, B. O. (1998–2021). Mplus user's guide (8th ed.). Los Angeles: Muthén & Muthén.

National Institute of Child Health, Human Development 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.

Ostrov, J. M., & Bishop, C. M. (2008). Preschoolers' aggression and parent–child conflict: A multi-informant and multimethod study. *Journal of Experimental Child Psychology, 99*, 309–322.

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.

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., & Keating, C. F. (2004). Gender differences in preschool aggression during free play and structured interactions: An observational study. *Social Development, 13*, 255–277.

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.

Pakaslahti, L., & Keltikangas-Jarvinen, L. (2000). Comparison of peer, teacher, and self-assessments on adolescent direct and indirect aggression. *Educational Psychology, 20*, 177–190.

Pellegrini, A. D. (1989). Elementary school children's rough-and-tumble play. *Early Childhood Research Quarterly, 4*, 245–260.

Pellegrini, A. D., & Bartini, M. (2000). An empirical comparison of methods of sampling aggression and victimization in school settings. *Journal of Educational Psychology, 92*, 360–366.

Perry, K. J., & Ostrov, J. M. (2018). Testing a bifactor model of relational and physical aggression in early childhood. *Journal of Psychopathology and Behavioral Assessment, 40*, 93–106.

Phares, V., Compas, B. E., & Howell, D. C. (1989). Perspectives on child behavior problems: Comparisons of children's self reports with parent and teacher reports. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 68–71.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*, 243–248.

Shin, H. J., Rabe-Hesketh, S., & Wilson, M. (2019). Trifactor models for multiple-ratings data. *Multivariate Behavioral Research, 54*, 360–381.

Sijtsema, J. J., & Ojanen, T. J. (2018). Social networks and aggression. In T. Malti & K. H. Rubin (Eds.), *Handbook of child and adolescent aggression* (pp. 230–248). New York: Guilford.

Swit, C. S., McMaugh, A. L., & Warburton, W. A. (2018). Teacher and parent perceptions of relational and physical aggression during early childhood. *Journal of Child and Family Studies, 27*, 118–130.

Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, O. D., & Boivin, M. (2005). Physical aggression during early childhood: Trajectories and predictors. *Canadian Child and Adolescent Psychiatry Review, 14*, 3–9.

Werner, N. E., Senich, S., & Przepyszny, K. A. (2006). Mothers' responses to preschoolers' relational and physical aggression. *Journal of Applied Developmental Psychology, 27*, 193–208.

Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2016). Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. *Research on Social Work Practice, 14*, 191–200.