Journal of Educational Psychology
2011, Vol. 103, No. 3, 701–732

© 2011 American Psychological Association
0022-0663/11/$12.00 DOI: 10.1037/a0024122

Construct Validity of the Multidimensional Structure of Bullying and Victimization: An Application of Exploratory Structural Equation Modeling

Herbert W. Marsh
University of Oxford, University of Western Sydney, and King Saud University

Benjamin Nagengast
University of Oxford

Alexandre J. S. Morin
University of Sherbrooke and University of Western Sydney

Roberto H. Parada
University of Western Sydney and University of Notre Dame Australia

Rhonda G. Craven and Linda R. Hamilton
University of Western Sydney

Existing research posits multiple dimensions of bullying and victimization but has not identified well-differentiated facets of these constructs that meet standards of good measurement: goodness of fit, measurement invariance, lack of differential item functioning, and well-differentiated factors that are not so highly correlated as to detract from their discriminant validity and substantive usefulness in school settings. Here we demonstrate exploratory structural equation modeling, an integration of confirmatory factor analysis and exploratory factor analysis. On the basis of responses to the 6-factor Adolescent Peer Relations Instrument (verbal, social, physical facets of bullying and victimization), we tested invariance of factor loadings, factor variances–covariances, item uniquenesses, item intercepts (a lack of differential item functioning), and latent means across gender, year in school, and time. Using a combination of relations with student characteristics and a multitrait–multimethod analysis, we showed that the 6 bully/victim factors have discriminant validity over time and in relation to gender, year in school, and relevant psychosocial correlates (e.g., depression, 11 components of academic and nonacademic self-concept, locus of control, attitudes toward bullies and victims). However, bullies and victims are similar in many ways, and longitudinal panel models of the positive correlations between bully and victim factors suggest reciprocal effects such that each is a cause and an effect of the other.

Keywords: exploratory structural equation modeling (ESEM) and measurement invariance, multiple dimensions of self-concept, depression, coping, locus of control

The present investigation is a substantive-methodological synergy (Marsh & Hau, 2007). It brings to bear new, strong, and evolving methodology in order to evaluate the psychometric properties of a new multidimensional instrument designed to measure bullying and victimization and to tackle complex substantive issues in this area of research. Although bullying and victimization are widely claimed to be multidimensional constructs, previous attempts to measure a priori factors based on a strong measure have met with only limited success. Here, as in other areas of research in psychology, theory, good measurement, research, and practice are inexorably related such that the neglect of one will undermine pursuit of the others. A particularly important starting point is the development of a solid measure of bullying and victimization, based on sound theory and supported by empirical research demonstrating support for convergent and discriminant validity. In this article, we begin with a review of substantive

This article was published Online First July 11, 2011.

Herbert W. Marsh, Department of Education, University of Oxford, Oxford, England; Department of Educational Psychology, University of Western Sydney, Penrith, New South Wales, Australia; and Department of Education, King Saud University, Riyadh, Saudi Arabia; Benjamin Nagengast, Department of Education, University of Oxford, England; Alexandre J. S. Morin, Department of Psychology, University of Sherbrooke, Sherbrooke, Québec, Canada, and Educational Excellence and Equity (E3) Research Program, Centre for Educational Research, University of Western Sydney; Roberto H. Parada, Educational Excellence and Equity (E3) Research Program, Centre for Educational Research, University of Western Sydney, and School of Arts and Sciences, University of Notre Dame Australia; Rhonda Craven and Linda R. Hamilton, Educational Excellence and Equity (E3) Research Program, Centre for Educational Research, University of Western Sydney.

This research was supported in part by a grant from the Australian Research Council. The authors would like to thank Linda Muthén for advice on analyses. Correspondence concerning the Adolescent Peer Relations Instrument should be sent to Roberto Parada, who can be reached at School of Arts and Sciences, University of Notre Dame Australia, 104 Broadway (P.O. Box 944), New South Wales 2007, Australia; roberto.parada@nd.edu.au.

Correspondence concerning this article should be addressed to Herbert W. Marsh, Department of Education, University of Oxford. E-mail: herb.marsh@education.ox.ac.uk

issues relevant to our study, introduce a new methodological approach, and summarize the present investigation.

Substantive Issue: Bullying, Victimization, and Psychological Correlates and Implications

Bullying is a growing and significant problem in many schools around the world with serious implications for the mental health and well-being of victims of bullying as well as bullies themselves (Due et al., 2005; Healey, 2001; Marsh, Parada, Yeung, & Healey, 2001; Mayer & Cornell, 2010; Olweus, 1991, 1993; Swearer, Espelage, Vaillancourt, & Hymel, 2010). Bullying involves an intentional, usually recurrent, action designed to inflict physical and psychological harm on another person or persons by one or more persons and is part of a complex interplay of dominance and social status (Sutton, Smith, & Swettenham, 1999). It incorporates a wide range of aggressive and social behaviors such as name calling, extortion, physical violence, slander, group exclusion, damage to property, and verbal intimidation (Smith & Sharp, 1994). Bullying is also different from aggression. Whereas aggression may involve a singular action by individuals or groups against each other and tends to be time limited (Feshbach & Zagrodzka, 1997), bullying is typically repetitive in nature, whereby bullies continue to bully victims for extended periods of time (Rigby, 1996). Bullying also involves a power imbalance between bullies and victims where the victims are often unable to defend themselves from the bullies (Olweus, 1997). This imbalance of power might, but does not necessarily, include physical superiority over the victim (e.g., Lagerspetz, Björqvist, Berts, & King, 1982) and can also be brought about by group membership or affiliation (i.e. the bully may belong to a gang, an exclusive clique, or a group with a racial or ethnic composition different to that of the victim). An imbalance can be related to specific attitudes and beliefs held by victims. For example, victims may be scared of the bully or may hold beliefs that prevent them from resorting to violence or from seeking social support to defend themselves (Parada, 2002). The key factor is that bullies, for their own benefit, exploit this imbalance of power to dominate victims repeatedly in an unwelcome way, resulting in victims being harmed or disadvantaged because of it. Bullies proactively seek victims due to characteristics that the perpetrator assesses as making the selected victim an easier target than others around them (Parada, 2006). These include knowledge that the bullies will be able to overpower victims, that the victims are unlikely to stand up for themselves or retaliate, and a context in which the bullies hold more power (e.g., if they have friends around to support them). In this sense, the victim is a target.

The Structure of Bullying and Victimization

In research on gender and developmental differences in children's aggressiveness (e.g., Björkqvist, Lagerspetz, & Kaukiainen, 1992), two forms of bullying have been differentiated: direct and indirect bullying, the latter also referred to as relational or social bullying (Gumpel, 2008; Rivers & Smith, 1994; Underwood, 2003). Direct bullying is characterized by behaviors that involve hitting, kicking, pinching, taking money or belongings, name calling, teasing, taunting, and threatening (Wolke, Woods, Bloomfield, & Karstadt, 2000). Crick and Grotpeter (1995) defined indirect bullying as mainly relational, being characterized by the hurtful manipulation of peer relationships and friendships to inflict harm on others through behaviors such as social exclusion and rumor spreading. The direct or indirect nature of bullying is therefore primarily defined by which method is used to bully the victim, where direct methods are characterized by overt behaviors (e.g., verbal and physical bullying). On the basis of these dimensions, the authors of many studies also have suggested that direct/indirect bullying may be further differentiated into, at least, physical, verbal, and antisocial/relational behaviors (e.g., Björkqvist et al. 1992; Crick et al., 2001; Finkelhor, Ormrod, & Turner, 2007; Mynard & Joseph, 2000; Rigby & Slee, 1999; Salmivalli, Kaukiainen, & Lagerspetz, 2000). Gumpel (2008) further noted the need to assess bullying and victimization simultaneously to better understand the complex patterns of relations between these domains and subdomains at a single point in time and longitudinally. While there is a clear theoretical rationale for this typology and many ad hoc instruments have been said to measure factors related to these subdomains, it seems that no such instrument has been shown to be psychometrically sound in relation to criteria considered in the present investigation (i.e., support for an a priori factor structure in relation the current standards of goodness of fit; construct validation in relation to convergent and discriminant validity; tests of differential item functioning in relation to gender, year in school, and over time; see discussion by Gumpel, 2008; Parada, 2006).

Aggressive behavior is not the same construct as bullying (see earlier discussion about the distinction), but the two constructs are related, and bullying often involves some form of aggression. In a recent meta-analysis of aggressive behavior, Card, Stucky, Sawalani, and Little (2008) found that the average correlation between direct (verbal and physical) and indirect aggression was a very high .76. Furthermore, they also found that correlations were even higher for boys but suggested that some measures might have blurred the differentiation between these constructs. The size of this correlation raises questions about the ability to distinguish between these two factors and whether the distinction is meaningful. At best, the size of the correlation obscures the nature of relations with psychosocial adjustment. Nevertheless, Card et al. did report some support for the discriminant validity of this distinction in that direct aggression was more strongly correlated with externalizing problems, poor peer relations, and low prosocial behaviors while indirect aggression was more strongly related to internalizing problems and less negatively related to prosocial behavior. It is interesting that these patterns of relations did not vary systematically with gender or age in meta-analysis results, leading the authors to argue against contention that "gender-nonnormative" forms of aggression are related to greater maladjustment.

In summary, there is some support for the ability to distinguish between direct and indirect forms of aggression, even though the correlations between these two constructs are extremely high. This suggests, perhaps, that there should be a parallel distinction between related forms of bullying but provides less clear guidance for whether a related distinction applies to victimization. Indeed, Card and Hodges (2008) concluded that "[u]nfortunately, a well-established, standardized measure of peer victimization does not exist, and commonly used scales assessing broad adjustment contain few items assessing victimization" (p. 452). In their recommendations for the development of assessment tools, they specif-

ically noted the need to differentiate among physical, verbal, and social–relational forms of victimization. Although there is some rigorous psychometric support for the separation of the broad bullying and victimization domains (e.g., Hussein, 2010), current best psychometric practices have not been applied extensively in this area of research.

Gender Differences in Bully and Victim Factors

Gender differences have been observed in school bullying. An extensive literature suggests that boys are more likely than girls to be bullies as well as victims (Boulton & Smith, 1994; Boulton & Underwood, 1992; Nansel et al. 2001). Girls, however, may engage in more covert forms of indirect bullying behaviors such as spreading rumors and encouraging social rejection and exclusion (Björkqvist, et al., 1992; Crick, Bigbee, & Howes, 1996; Crick, & Grotpeter, 1995) so that the actual rates of bullying/victimization attributed to girls may have been underestimated. In a recent review of gender differences in aggressive behavior, Ostrov and Godleski (2010) concluded that based on the results from two previous meta-analyses (Archer, 2004; Card et al., 2008), there is at best weak support for the gender-linked hypothesis that males engage in more physical aggression while females engage in more indirect social–relational aggression. Indeed, in a meta-analysis conducted by Archer (2004) based on self-reports, the sex differences were slightly smaller for verbal than for physical aggression, but there were no significant gender differences in indirect aggression. However, Archer qualified this result, noting that some measures of indirect aggression used in the studies in the meta-analysis included a diverse mix of items so that gender differences might be confounded with the content of different items. Card et al. (2008) also found, based on self-reports, large differences favoring boys in direct aggression but only "trivial" (p. 1185) differences in indirect aggression, again slightly in favor of boys. Why, they asked, have the popular media and researchers themselves argued that girls have higher levels of indirect aggression than boys? In response, they suggested that popular accounts of gender differences in forms of aggressive behaviors distort perceptions of parents and even teachers—but apparently not self-reports by students themselves.

In summary, there is clear support for more aggressive behavior among boys than girls for direct forms of aggression. While gender differences are smaller for indirect forms, there appears to be no consistent support for the claim that girls engage more in indirect forms of aggression than boys. As noted previously, aggressive behavior is not necessarily the same construct as bullying, and results based on aggressive behavior do not necessarily provide much information about victimization. Hence, it is important to test the gender-linked hypothesis with psychometrically sound measures of different subdomains of bully and victim factors. Indeed, there is little rigorous psychometric research demonstrating the invariance of factor structures over gender that is an implicit, typically untested assumption of tests of mean level differences across gender. If bully and victim factors are qualitatively different for boys and girls, then comparisons of mean level differences are likely to be meaningless (the apples and oranges phenomena). Nevertheless, establishing whether gender differences exist in bully and victim domains as well as in subdomains of these constructs is of great practical significance to understanding these constructs and to the development of intervention programs.

Who Are the Victims and the Bullies? Rethinking Bipolar Classification Schemes and Dichotomizing Bully and Victim Variables

The measurement of bullying and victimization has been dominated by use of simple surveys, often based on a single item to define the various bullying constructs such as direct, indirect, verbal, or physical bullying and victimization (e.g., Rigby, 1996; Smith & Sharp, 1994). From their review of research on the relations between victimization and mental health published between 1980 and 2000, Hawker and Boulton (2000) concluded that the complexity of the bullying phenomenon has been largely ignored in measuring bullying. In studies of the link between victimization and depression, for example, "only one study measured relational or indirect victimization and this study did not include verbal victimization [and] only one study used more than one item to assess peer-reported victimization" (p. 447). Bosworth, Espelage, and Simon (1999) also highlighted several shortcomings of international efforts to measure bullying/victimization in schools such as (a) studies that have concentrated on the extreme ends of what may very well be a continuum of bullying behaviors and (b) surveys in which students were presented with a definition of bullying prior to their being asked to self-report their engagement in such behavior, which might increase social desirability bias. They therefore advocated a continuum approach to the measurement of bully and victim factors and suggested that it may be more appropriate to "simply ask students about the frequency of specific behaviors such as teasing and hitting" (Bosworth et al., 1999: p. 343).

Problems in the classification and measurement of bullying and being bullied are due in part to historical tendencies to classify students as either bullies or victims and to implicit assumptions that they represent the endpoints of a bipolar continuum. We now know that this is inappropriate. A growing body of research from around the world (Harachi, Catalano, & Hawkins, 1999; Roland & Idsøe, 2001; Smith et al., 1999; Sullivan, 2000) shows that bullying and victimization tend to be positively correlated. Whereas there is an ongoing debate about the size, nature, and underlying processes associated with this correlation, the correlation is clearly not the -1.0 correlation that would be consistent with the mutually exclusive or bipolar classification schemes with which some researchers devise dubious cutoff values to classify students. Although the practice of dichotomizing continuous scores has been broadly recognized as largely inappropriate in psychological research for more than a quarter of a century (Cohen, 1968; Cohen & Cohen, 1983) and is diminishing in most areas of applied psychological research, the practice is still widespread in the bullying literature. While this tradition might be useful for clinical practice, it is counterproductive to research (see review by MacCallum, Zhang, Preacher, & Rucker, 2002, documenting the few instances when it is appropriate).

Few attempts have been made to explain why bully and victim factors are positively correlated. A logical explanation is that rather than being mutually exclusive roles, bullying and victimization may be mutually reinforcing, whereby prior bullying behavior may lead to later victimization, and prior victimization may

lead to later bullying behavior (Card & Hodges, 2008; Marsh, Parada, Craven, & Finger, 2004; Marsh et al., 2001; Ostrov, in press; Parada, 2000, 2006; Parada, Marsh & Craven, 2005). Viewing bullying and victimization as mutually reinforcing constructs and explaining the causal relations between bully and victim factors might explain how certain characteristics are more likely to instigate bullying and victimization behaviors. The explanation can also shed light on how bullying and victimization are related—negatively or positively—to psychological factors, specifically depression, locus of control, anger management, self-concept, and psychosocial adjustment. Testing causal hypotheses of this type requires, at a minimum, longitudinal models in which the same construct is measured on multiple occasions (Marsh & Craven, 2006).

This hypothesized pattern of longitudinal relations between bully and victim factors also suggests the importance of evaluating these factors as a function of age or year in school. Due to an inherent power imbalance, the youngest students in a school are more likely to victims of older students. However, as these students grow older, it is easier for them to bully younger students.

Methodological Focus: Harnessing the Power, Flexibility, and Versatility of Exploratory Structural Equation Modeling (ESEM)

Many psychological instruments have an apparently well-defined exploratory factor analysis (EFA) structure but cannot be represented adequately within a confirmatory factor analysis (CFA) approach (Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010). Typically, this is the result of their factor structures not being consistent with the highly restrictive independent clusters model (ICM) typically used in CFA studies in which each item is allowed to load on only one factor, and nontarget loadings are constrained to be zero. Although there are many methodological and strategic advantages to ICM CFA, they are sometimes inappropriate, and many of the strategies traditionally used to compensate for this inappropriateness (e.g., use of item parcels; Marsh et al., 2011) tend to be dubious, counterproductive, misleading, or simply wrong. Furthermore, the misspecification of zero factor loadings usually leads to distorted factors with overestimated factor correlations that might lead to biased estimates in structural equation models (SEMs) incorporating other outcome variables (Asparouhov & Muthén, 2009; Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010). Indeed, elevated correlations between subscales designed to measure bullying appear to be the norm rather than the exception (Card et al., 2008). The few strategies that have been proposed to circumvent this potential problem involve highly complex factor structures that do not manage to satisfactorily reduce all of the observed correlations (e.g., Card & Little, 2007; Little, Jones, Henrich, & Hawley, 2003). In related research, Marsh (2007a; Marsh, Hau, & Grayson, 2005) argued that few multidimensional assessment instruments meet even minimal standards of goodness of fit based on CFA. Such research led Marsh, Hau, and Wen (2004) to question the appropriateness of treating rough guidelines about goodness of fit as if they were "golden rules."

Here we describe an application of exploratory structural equation modeling (ESEM), an evolving, new statistical procedure in which many of the best features of EFA, CFA, and SEM are integrated. Within the ESEM framework, the applied researcher has access to typical SEM parameters, standard errors, goodness-of-fit statistics, and many statistical advances normally associated with CFA and SEMs: systematic comparisons of competing models through tests of statistical significance and fit indices; inclusion of correlated residuals; estimation of method effects and bifactor models; estimation of path and multiple indicators multiple causes models (MIMIC) models relating covariates to latent factors; multiple group invariance analyses, including full mean structures and tests of differential item functioning; tests of invariance over time for multiwave data; growth modeling (for illustrations of a diverse range of ESEM models, see Asparouhov & Muthén, 2009; Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010). In the present investigation, we demonstrate many of these new possibilities and the new substantive insights they offer on the basis of responses to the Adolescent Peer Relations Instrument (APRI).

Of particular substantive importance for educational psychology studies is the examination of mean-level differences across multiple groups (e.g., male vs. female; year in school groups; clinical vs. nonclinical participants; experimental vs. control groups) or over time (i.e., observing the same group of participants at multiple occasions, perhaps before and after an intervention). What has typically been ignored in educational psychology studies are tests of whether the underlying factor structure is the same in the different groups or on multiple occasions (Card, & Little, 2007; Selig, Card, & Little, 2008). Unless this evidence is present, mean differences and other comparisons are likely to be invalid. If the underlying factor is qualitatively different for boys and girls, then mean differences are not justified and uninterpretable—the apples and oranges phenomena. However, traditional EFA approaches are not suitable for testing invariance and ICM–CFA approaches typically are not able to provide an acceptable fit to the data when applied to complex multidimensional instruments—even when no invariance constraints are imposed. Hence, tests of invariance have rarely been considered in bullying research (but see Vaillancourt, Brendgen, Boivin, & Tremblay, 2003), even though such assumptions are implicit in most studies (Card & Little, 2007).

In summary, ESEM offers important methodological contributions to all aspects of bullying research that are particularly relevant to the present investigation. Even when ICM–CFA models provide a moderately acceptable fit to the data, the relations among the factors are likely to be inflated. This undermines the ability to differentiate between different bully and victim factors and distorts their relations with other constructs. Mean level differences between different groups or comparisons over time for the same group are based on assumptions of invariance that are rarely considered in bullying research, but these tests are easily implemented within an ESEM application.

The Present Investigation

Substantively, the present investigation is part of a research program designed to better understand bullying and victimization in high schools. Our substantive aims are to address issues relevant to understanding bullying and victimization by (a) testing the psychometric properties and measurement invariance of an instrument specifically designed to measure multiple bully and victim factors (the APRI; Parada, 2000; also see Marsh, Parada, et al., 2004; Parada, Craven, & Marsh, 2008; also see Appendices A and

B that contain the instrument and scoring key); (b) testing the effects of year in school and gender in relation to bully and victim factors; (c) evaluating the relation between bully and victim factors; (d) evaluating relations between bully and victim factors and a variety of psychosocial correlates (depression, 11 components of academic and nonacademic self-concept, locus of control, attitudes towards bullies and victims; see Appendix A), and (e) using ESEMs of longitudinal data to disentangle the directionality of the processes underlying the positive correlations between bully and victim factors. The present investigation was a substantive-methodological synergy that explored the utility of ESEM as a technique integrating many of the advantages of CFA, SEM, and EFA. In order to facilitate the presentation, we have divided the article into Study 1 (cross-sectional analyses based only on responses from the first wave) and Study 2 (longitudinal analyses based on responses from all three waves).

Study 1

We begin our analyses with a detailed evaluation of responses to the bully and victim subscales on the basis of cross-sectional analyses of a single wave of data. Our purposes in Study 1 were to evaluate the a priori factor structure of the APRI measures of bully and victim factors, their measurement invariance in relation to gender and year in school, and their convergent and discriminant validity in relation to a variety of relevant constructs. Consistent with our substantive-methodological focus, we validate the a priori six-factor structure and demonstrate the superiority of the ESEM approach in comparison to the traditional ICM–CFA approach. We then pursue a series of tests of measurement invariance of the bully/victim factor structure over gender and year in school. Finally, we evaluate support for the convergent and discriminant validity of the bully/victim factors in relation to a purposively chosen set of psychosocial constructs with substantive implications for understanding the nature of bullying and victimization.

Method

Participants. The eight schools participating in the present investigation were drawn from high schools affiliated with a large Catholic education office diocese west of Sydney. Participants (N = 4,082 students) were in Years 7–11 (M age = 13.8, SD = 1.4; 42.9% boys, 86.1% Australian born). The questionnaire package was administered to all students on three occasions during one school year, although the response rates varied somewhat for each occasion: Time 1 (T1) early in the first school term (March)—87% response rate (n = 3,512); Time 2 (T2) at the end of term 2 (July)—87% response rate (n = 3,557), and Time 3 (T3) at the end of term 4 (December)—80% response rate (n = 3,263). Within each wave, there were nearly complete data for those who were in attendance on the day the data were collected (less than 1% missing).

Adolescent Peer Relations Instruments (APRI). The APRI (see Appendix A) comprises multidimensional measures of bully/victim behaviors specifically developed for this study (Marsh, Parada, et al., 2004; Parada, 2000). The two domains (bully and victimization) are each assessed in relation to three subdomains (verbal, physical, and social), resulting in a total of six scales. Students were asked to state how often in the past year, on a 6-point Likert scale (from 1 = *never* to 6 = *everyday*), they had engaged in a series of behaviors against other students. The second section asked how often these behaviors had occurred to them. There are a total of 36 items, six for each of the three bullying scales and six for each of the three victimization scales. Sample items and coefficient alpha based on each of three waves of data are as follows: (a) Bully–verbal: Teased them by saying things to them (α = .89, .90, .92); (b) Bully–social: Got my friends to turn against a student (α = .82, .86, .90); (c) Bully–physical: Got into a physical fight with a student because I didn't like him or her (α = .85, .87, .90); (d) Victim–verbal: I was teased by students saying things to me (α = .92, .92, .93); (3) Victim–social: A student wouldn't be friends with me because other people didn't like me (α = .87, .91, .92); (f) Victim–physical: I was threatened to be physically hurt or harmed (α = .89, .89, .92).

As part of the analysis, we related the bully and victim factors to a variety of psychosocial correlates: depression, 11 components of academic and nonacademic self-concept, locus of control, attitudes toward bullies and victims, and participant roles (these constructs and basic psychometric information are presented in the Appendix B).

Statistical analyses. All analyses in the present investigation were done with Mplus 5.2 (Muthén & Muthén, 2008). Preliminary analyses consisted of a traditional CFA model based on the Mplus robust maximum likelihood estimator (MLR) with standard errors and tests of fit that are robust in relation to nonnormality (Muthén & Muthén, 2008). A main focus is on the application of ESEM and its comparison to traditional CFA solutions. The ESEM approach differs from the typical CFA approach in that all factor loadings are estimated, subject to identification constraints (for further details of the ESEM approach and identification issues, see Asparouhov & Muthén, 2009; Marsh et al., 2009). In ESEMs, items within the same sets of factors are free to cross-load on other factors within the same set of factors, whereas in CFAs all cross-loadings are specified to be zero. Because some of the ICM–CFA factors were highly correlated at .7 or more (Parada, 2006; also see subsequent presentation of results from the current study) and following recommendations by Marsh, Lüdtke, Muthén, et al. (2010) for similar situations, we used an oblique geomin rotation with an epsilon value of .5. We relied on the full-information MLR estimator (see Enders, 2001; Graham, 2009) to account for missing data. More detailed descriptions of alternative rotation procedures are available elsewhere (Asparouhov & Muthén, 2009; Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010; Sass & Schmitt, 2010).

Multiple group analysis. The evaluation of model invariance over different groups (e.g., gender) is widely applied in SEM studies (Jöreskog & Sörbom, 1979; Meredith, 1993; Meredith & Teresi, 2006). Indeed, such tests of invariance might be seen as a fundamental advantage of CFA–SEM over EFA. Although related multiple group methods have been proposed in EFA settings (e.g., Cliff, 1966; Meredith, 1964), they mainly focus on the similarity of factor patterns (but also see Dolan, Oort, Stoel, & Wicherts, 2009). However, the ESEM model can be extended to multiple group analyses, such that the ESEM solution is estimated separately for each group and some or all parameters can be constrained to invariance across groups.

To test factorial and measurement invariance across multiple groups, Marsh et al. (2009; Marsh, Lüdtke, Muthén, et al., 2010)

operationalized a taxonomy of 13 partially nested models varying from the least restrictive model of configural invariance with no invariance constraints to a model of complete invariance positing strict invariance as well as the invariance of the latent means and of the factor variance–covariance matrix (Table 1). This taxonomy begins with a model with no invariance of any parameters or *configural invariance* (Model 1). The initial focus is on the invariance of the factor loadings (Model 2)—sometimes referred to as weak measurement invariance or pattern invariance—which requires that factor loadings be invariant over groups. *Strong measurement invariance* (Model 5) requires that the indicator intercepts and factor loadings are invariant over groups and is a prerequisite to the comparison of latent means. *Strict measurement invariance* (Model 7) requires invariance of item uniquenesses (addition to invariant factor loadings and intercepts), a necessary condition for the comparison of manifest means over groups. However, because we are evaluating latent models and not models of manifest scores, Model 7 is not essential, while Model 5 and the issue of differential item functioning are critical.

In order to evaluate the construct validity of the APRI responses, we then related the bully/victim factors to a diverse set of other psychosocial variables specifically designed to address key questions about the nature of bullying (see Appendices A and B). This ESEM model was very large, positing 32 latent factors and four (single-indicator) demographic variables (representing gender and year in school) based on responses to a total of 172 items. Further demonstrating the flexibility of the ESEM approach, the 32-factor model contained 10 sets of ESEM factors (three for bully, three victim, four participant role, three coping style, two attitudes toward bullying and victimization, two locus of control, three anger management, one depression, eight nonacademic self-concept, and three academic self-concept; see Appendix A for further discussion of these factors) as well as the demographic variables. Thus, within each set of factors, items were allowed to cross-load on different factors but were not allowed to cross-load on factors from different sets. Finally, we sought to examine how the bully/victim factors and variables directly related to them (attitudes toward bullies and victims and participant roles in bullying situations) varied as a function of gender, year in school, and their interaction. We expected that the bully/victim domain scores would be higher for boys than girls but that gender differences were specific to the subdomains (verbal, social, physical). Thus, based on previous research, we posited that boys would have higher scores on the physical and, perhaps, verbal subdomains, while gender differences should be small for the social subdomain. To this end, each of these factors was regressed on four contrast variables: gender (male vs. female), year (linear and quadratic), and Year × Gender interaction (as the Quadratic Year × Gender interaction was not significant here, and in subsequent analyses, this component was not considered further).

Table 1
Taxonomy of Multiple Group Tests of Invariance Testable With Exploratory Structural Equation Modeling

Model	Parameters constrained to be invariant
1	None (configural invariance)
2	FL [1] (weak factorial/measurement invariance)
3	FL Uniq [1, 2]
4	FL, FVCV [1, 2]
5	FL, Inter [1, 2] (strong factorial/measurement invariance)
6	FL, Uniq, FVCV [1, 2, 3, 4]
7	FL, Uniq, Inter [1, 2, 3, 5] (strict factorial/measurement invariance)
8	FL, FVCV, Inter [1, 2, 4, 5]
9	FL, Uniq, FVCV, Inter [1–8]
10	FL, Inter, FMn [1, 2, 5] (latent mean invariance)
11	FL, Uniq, Inter, FMn [1, 2, 3, 5, 7, 10] (manifest mean invariance)
12	FL, FVCV, Inter, FMn [1, 2, 4, 5, 6, 8, 10]
13	FL, Uniq, FVCV, Inter, FMn [1–12] (complete factorial invariance)

Note. Models with latent factor means freely estimated constrain intercepts to be invariant across groups, while models where intercepts are free imply that mean differences are a function of intercept differences. Brackets values represent nesting relations in which the estimated parameters of the less general model are a subset of the parameters estimated in the more general model under which it is nested. All models are nested under Model 1 (with no invariance constraints) while Model 13 (complete invariance) is nested under all other models. FL = factor loadings; FVCV = factor variance–covariances; Inter = item intercepts; Uniq = item uniquenesses; FMn = factor means.

Goodness of fit. In applied CFA–SEM research, there is a predominant focus on indices that are sample-size independent (e.g., Marsh, 2007a; Marsh, Balla & Hau, 1996; Marsh, Balla, & McDonald, 1988; Marsh, Hau & Grayson, 2005) such as the root-mean-square error of approximation (RMSEA), the Tucker–Lewis Index (TLI), and the confirmatory fit index (CFI). Thus, for consistency with previous work, these three indices routinely provided by Mplus (Muthén & Muthén, 2008) will be reported, as well as the MLR chi-square test statistic and an evaluation of parameter estimates. Both, TLI and CFI vary along a 0-to-1 continuum and values greater than .90 and .95 typically reflect acceptable and excellent fit to the data. RMSEA values of less than .05 and .08 reflect a close fit and a reasonable fit to the data, respectively (Marsh, Hau, & Wen, 2004). However, for purposes of model comparison, comparison of the relative fit of models imposing more or fewer invariance constraints is more important than the absolute level of fit for any one model—so long as the fit of the best-fitting model is acceptable. Cheung and Rensvold (2002) and Chen (2007) suggested that if the decrease in fit for the more parsimonious model is less than .01 for incremental fit indices like the CFI, then there is reasonable support for the more parsimonious model. Chen (2007) suggested that when the RMSEA increases by less than .015, there is support for the more constrained model. For indices that incorporate a penalty of lack of parsimony, it is also possible for a more restrictive model to result in a better fit than a less restrictive model. Although we relied on these guidelines in the present investigation, we should emphasize that these are only rough guidelines rather than golden rules (Marsh, Hau & Wen, 2004) so that ultimately there is a degree of subjectivity and professional judgment required in the selection of a "best" model (Marsh et al., 1988). Furthermore, additional research is needed concerning their appropriateness for ESEM studies, for which the number of estimated parameters is typically substantially greater than the typical ICM–CFA study. In the meantime, we suggest that applied researchers use an eclectic approach based on a subjective integration of a variety of different indices—including the chi-square, detailed evaluations of the actual parameter estimates in relation to theory, a priori predictions,

common sense, and a comparison of viable alternative models specifically designed to evaluate goodness of fit in relation to key issues. This is consistent with the approach we used here.

Results

APRI Factor Structures and Correlations Among Bully/Victim Factors: ESEM vs. CFA. The critical starting point for the present investigation was to test the hypothesis that the a priori ESEM model provides a better fit to responses to the two sets of 18 APRI items than a traditional ICM–CFA model and that it reduces typically high and problematic interfactor correlations. The ICM–CFA solution provides an acceptable fit to the data (CFI = .943, TLI = .938, RMSEA = .029; see total-group (TG) CFA in Table 2). However, the corresponding ESEM solution fits the data significantly better (CFI = .963, TLI = .955, RMSEA = .025; see TG–ESEM in Table 2). Hence, there is support for the ESEM approach over the traditional ICM–CFA approach, but the difference is not as substantial as has been found in other applications of ESEM (e.g., Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010). Hence, the question now becomes whether other features of ESEM justify its application over the more parsimonious CFA.

In terms of the 36 target factor loadings (see Table 3), the sizes of most loadings are substantial for both the ESEM (.425–.736; M = .592) and the CFA (.596–.848) solutions. For the ESEM solution, ESEM nontarget loadings are systematically smaller (from −.066 to .325, M =.126) than ESEM target loadings (nontarget loadings are constrained to be zero in the CFA solution). When both target and nontarget factor loadings are considered together, the ICM–CFA and ESEM solutions result in a very similar pattern, with a profile similarity index (PSI; the correlation between the set of ESEM factor loadings and the corresponding CFA factor loadings) of .935, suggesting that the ESEM and ICM–CFA factor loadings were highly related.

An evaluation of the factor correlations shows a critical advantage of the ESEM approach over the ICM–CFA approach. Although patterns of correlations between the ESEM and CFA solutions are similar (PSI = .948), the ICM–CFA factor correlations (.128–.836; median [Md] = .589) are substantially higher than the ESEM factor correlations (.011–.527; Md = .274). Indeed, inspection of Table 3 shows that correlations among the three bully factors and among the three victim factors range from .72 to .84 for the CFA solution, but only from .42 to .53 for the ESEM solution. These findings are supportive of the claim by Marsh and colleagues (Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010; also see Asparouhov & Muthén, 2009) that the sizes of factor correlations in ICM–CFA tend to be systematically inflated compared with those in ESEM. This systematic bias varies with the size of the ESEM cross-loadings (which are constrained to zero in the CFA solution). In this respect, the six bully–victim factors are substantially more distinct for the ESEM solution than for the CFA solution. Particularly for purposes of individual diagnosis and better understanding the nature of bullying, the very high CFA correlations detract substantially from the usefulness of the APRI but are typical of results from other research. From this perspective, the contribution of ESEM is substantively important.

Invariance across gender: Latent mean structure approach. The key questions we addressed here are: How consistent is the factor structure of bully and victim factors over gender? Are there systematic gender differences in latent means, and are the underlying assumptions met to justify interpretations of these results? The basic strategy in pursuing this goal is to apply the set of 13 models presented in Table 1.

Weak factorial/measurement invariance (multiple-group gender–Model 1 vs. multiple-group gender–Model 2; see Table 2). Weak factorial/measurement invariance tests whether the factor loadings are the same for girls and boys. In support of the invariance of factor loadings over gender, fit indices that control for parsimony are nearly the same for the more parsimonious multiple-group gender–Model 2 (MGG–M2) than the less parsimonious MGG–M1 (TLI = .952 vs. .951; RMSEA = .025 vs. .025). Thus, the differences in RMSEA and CFI (.976 vs. .977) are less than the respective .015 and .01 cutoff values typically used to reject the more parsimonious model. These results provide reasonable support for weak factorial/measurement invariance) of the ESEM factor structure over gender.

Strong measurement invariance* (*multiple-group gender–Model 5 vs. multiple-group gender–Model 2; see Table 2). Strong measurement invariance tests whether item intercepts—as well as factor loadings—are invariant over groups. Nonsupport for this model would imply differential item functioning (although it would still be possible to posit partial invariance; see Marsh et al., 2009; Marsh, Lüdtke, Muthén, et al., 2010), whereas strong support for this model would imply that differences between groups at the item level can be explained in terms of differences at the latent factor mean level. Fit indices for MGG–M5 are comparable to those of MGG–M2 (TLI = .951 vs. .950; CFI = .957 vs. .954; RMSEA = .026 vs. .025), demonstrating strong measurement invariance for bully/victim responses over gender and a justification for the comparison of latent mean differences.

Strict measurement invariance (multiple-group gender–Model 7 vs. multiple-group gender–Model 5; see Table 2). Strict measurement invariance requires that item uniqueness, item intercepts, and factor loadings are all invariant across groups. A lack of support for this model would suggest that measurement error differs in the two groups and thus preclude the comparison of manifest scale or factor scores of the bully/victim factors. Indeed, fit indices for MGG–M7 are systematically poorer than those for MGG–M5 (TLI = .855 vs. .950; CFI = .864 *vs.* .954; RMSEA = .040 *vs.* .026). Inspection of the results shows that measurement errors are systematically larger for boys than girls. While it might be possible to pursue a strategy of partial invariance of uniquenesses, the substantial differences in fit militate against the use of such a strategy. Because measurement error is corrected in ESEM models, this is not a necessary assumption for the valid comparison of latent means of boys and girls based on MGG–M5.

Factor variance–covariance invariance (multiple-group gender–Model 4 vs. multiple-group gender–Model 2; see Table 2). Factor variance–covariance invariance is typically not a focus in studies of measurement invariance but is frequently an important focus of studies of the discriminant validity of multidimensional constructs that might subsequently be extended to include relations with other constructs. While the comparison of correlations between factors for different groups such as gender is common, these are typically based on manifest scores that do not control for measurement error (particularly important if measurement error differs for the groups as it is the case here) and make implicit invariance assumptions that are rarely tested. However, our results do not support the

Table 2
Goodness-of-Fit Indices for Models of Wave 1 Data

Model	χ^2	*df*	CFI	TLI	No. FParms	RMSEA	Invariance constraints
				Total group analysis—Wave 1			
TG–CFA	2316	579	.943	.938	123	.029	None
TG–ESEM	1650	519	.963	.955	183	.025	None
				Multiple-group gender invariance—Wave 1			
MGG–M1	2212	1038	.960	.952	366	.025	None
MGG–M2	2414	1128	.957	.951	276	.025	FL
MGG–M3	4474	1164	.888	.879	240	.04	FL, Unq
MGG–M4	3053	1149	.936	.929	255	.031	FL, FVCV
MGG–M5	2519	1158	.954	.950	246	.026	FL, INT
MGG–M6	5213	1185	.864	.855	219	.044	FL, Unq, FVCV
MGG–M7	4604	1194	.885	.878	210	.04	FL, Unq, INT
MGG–M8	3161	1179	.933	.928	225	.031	FL, FVCV, INT
MGG–M9	5347	1215	.860	.855	189	.044	FL, Unq, FVCV, INT
MGG–M10	2803	1164	.945	.940	240	.028	FL, INT, FMn
MGG–M11	4879	1200	.876	.869	204	.042	FL, Unq, INT, FMn
MGG–M12	3505	1185	.922	.917	219	.033	FL, FVCV, INT, FMn
MGG–M13	5662	1221	.850	.845	183	.046	FL, Unq, FVCV, INT, FMn
				Multiple-group year invariance—Wave 1			
MGY–M1	4679	2595	.939	.926	915	.034	None
MGY–M2	4913	2955	.942	.939	555	.031	FL
MGY–M3	5863	3099	.919	.917	411	.036	FL, Unq
MGY–M4	5157	3039	.938	.935	471	.032	FL, FVCV
MGY–M5	5214	3075	.937	.936	435	.031	FL, INT
MGY–M6	6143	3183	.913	.914	327	.036	FL, Unq, FVCV
MGY–M7	6161	3219	.914	.915	291	.036	FL, Unq, INT
MGY–M8	5457	3159	.932	.933	351	.032	FL, FVCV, INT
MGY–M9	6443	3303	.908	.912	207	.037	FL, Unq, FVCV, INT
MGY–M10	5427	3099	.932	.930	411	.033	FL, INT, FMn
MGY–M11	6360	3243	.908	.911	267	.037	FL, Unq, INT, FMn
MGY–M12	5655	3183	.927	.928	327	.033	FL, FVCV, INT, FMn
MGY–M13	6624	3327	.903	.908	183	.038	FL, Unq, FVCV, INT, FMn
				Multiple-group time invariance—Waves 1–3			
MGT–M0	13561	5337	.941	.936	657	.019	None, no CWCUs
MGT–M1	10559	5229	.962	.958	765	.016	None
MGT–M2	10748	5409	.962	.959	585	.016	FLs
MGT–M3	11487	5481	.957	.955	513	.016	FLs & Unq
MGT–M4	11490	5451	.957	.954	543	.016	FLs & FVCV
MGT–M5	10900	5469	.961	.959	525	.016	FLs, INT
MGT–M6	12281	5523	.952	.949	471	.017	FL, Unq, FVCV
MGT–M7	11641	5541	.956	.954	453	.016	FL, Unq, INT
MGT–M8	11645	5511	.956	.954	483	.017	FL, FVCV, INT
MGT–M9	12438	5583	.951	.949	411	.017	FL, Unq, FVCV, INT
MGT–M10	11099	5481	.960	.958	513	.016	FL, INT, FMn
MGT–M11	11835	5553	.955	.953	441	.017	FL, Unq, INT, FMn
MGT–M12	11820	5523	.955	.953	471	.017	FL, FVCV, INT, FMn
MGT–M13	12609	5595	.950	.948	399	.018	FL, UNQ, FVCV, INT, FMn

Note. See Table 1 for a description of the models (M1–M13). CFI = comparative fit index; TLI = Tucker–Lewis index; No. FParm = number of free parameters; RMSEA = root-mean-square error of approximation; TG = total group; CFA = confirmatory factory analysis; ESEM = exploratory structural equation modeling; MMG = multiple-group gender; MGY = multiple-group year; MGT= multiple-group time. For multiple group invariance models, the parameters constrained to be invariant across the multiple groups (or multiple times for the longitudinal data) are FL = factor loadings; Unq = item uniquenesses; FVCV = factor variance–covariances; INT = item intercepts; FMn = factor means; CWCU = cross-wave-correlated uniquenesses.

imposition of these additional invariance constraints on the variance–covariance matrix, either in terms of the values for the fit indices or their comparison with MGG–M2. For example, fit indices for MGG–M4 are systematically poorer than for MGG–M2 (.929 *vs.* .951 for TLI; .031 *vs.* .025 for RMSEA; .936 *vs.* .957 for CFI). Other pairs of models that differ only with respect to the invariance of the latent factor variance–covariance matrix could also be used for this comparisons (as they could have been in the preceding comparisons) and result in similar conclusions (e.g., MGG–M2 *vs.* MGG–M4; MGG–M3 *vs.* MGG–M6; MGG–M5 *vs.* MGG–M8; MGG–M7 *vs.* MGG–M9; MGG–M10 *vs.* MGG–M2; MGG–M11 *vs.* MGG–M13).

Factor correlations (Table 4) among bully/victim factors show a substantially different pattern of results for boys and

Table 3
Comparison of Exploratory Structural Equation Modeling (ESEM) and Independent Clusters Model–Confirmatory Factory Analysis (ICM–CFA) Factor Structures

	ESEM factor solution								
	Factor loading								
	Bully			Victim				ICM–CFA solution	
Factor/model	Verbal	Social	Physical	Verbal	Social	Physical	R^2	Factor loading	R^2
Bully–verbal									
BV1	**.66**	.03	.14				.57	**.74**	.54
BV2	**.64**	.06	.21				.64	**.80**	.63
BV3	**.70**	.11	.04				.60	**.75**	.56
BV4	**.48**	.08	.31				.55	**.74**	.55
BV5	**.43**	.30	.14				.50	**.70**	.49
BV6	**.69**	.16	.07				.67	**.81**	.66
Bully–social									
BS1	.08	**.59**	.08				.46	**.67**	.45
BS2	.05	**.53**	.23				.48	**.70**	.49
BS3	.04	**.58**	.08				.40	**.62**	.38
BS4	.04	**.72**	.01				.55	**.70**	.50
BS5	.14	**.51**	.11				.41	**.66**	.43
BS6	.20	**.46**	.06				.37	**.61**	.38
Bully–physical									
BP1	.25	.02	**.56**				.53	**.72**	.53
BP2	.20	.19	**.45**				.48	**.71**	.50
BP3	−.06	.20	**.56**				.41	**.60**	.36
BP4	.09	.04	**.68**				.55	**.72**	.52
BP5	.19	.09	**.58**				.56	**.76**	.58
BP6	.09	.13	**.57**				.49	**.69**	.47
Victim–verbal									
VV1				**.74**	.00	.14	.66	**.79**	.61
VV2				**.65**	.10	.18	.67	**.81**	.65
VV3				**.67**	.20	.08	.71	**.84**	.70
VV4				**.59**	.20	.10	.61	**.79**	.62
VV5				**.46**	.28	.24	.66	**.81**	.66
VV6				**.72**	.14	.08	.72	**.85**	.72
Victim–social									
VS1				.19	**.54**	.09	.52	**.73**	.53
VS2				.16	**.60**	.01	.49	**.69**	.47
VS3				.16	**.68**	.05	.64	**.78**	.61
VS4				.06	**.64**	.12	.56	**.73**	.54
VS5				.24	**.42**	.26	.58	**.76**	.58
VS6				.02	**.56**	.28	.57	**.74**	.55
Victim–physical									
VP1				.22	−.04	**.66**	.59	**.74**	.55
VP2				.08	.06	**.71**	.61	**.76**	.58
VP3				.20	.11	**.57**	.58	**.77**	.59
VP4				.03	.33	**.52**	.58	**.75**	.57
VP5				.11	.15	**.62**	.60	**.78**	.60
VP6				.07	.23	**.57**	.57	**.75**	.57
	Factor correlations (ICM–CFA above diagonal, ESEM below)								
Bully–verbal	—	.72	.83	.27	.14	.27			
Bully–social	.42	—	.73	.17	.18	.22			
Bully–physical	.53	.44	—	.21	.13	.34			
Victim–verbal	.27	.08	.15	—	.84	.81			
Victim–social	.05	.17	.01	.43	—	.83			
Victim–physical	.22	.15	.38	.51	.52	—			

Note. The exploratory structural equation model (ESEM) solution was based on the 18 items from the bully instrument and 18 items from the victim instrument (see Model TG–ESEM in Table 2 for goodness-of-fit statistics). For both ESEM and independent clusters model–confirmatory factory analysis (ICM–CFA) solutions, all parameter estimates are standardized and a priori target loadings designed to measure each factor are in bold. In order to conserve space, we present the CFA factor loadings in condensed format such that only the target loading relating each item to its a priori factor is presented (as all nontarget loadings are zero). BV = bully–verbal; BS = bully–social; BP = bully–physical; VV = victim–verbal; VS = verbal–social; VP = verbal–physical.

Table 4
Correlation Matrix of Bullying and Victimization Factors for Boys and Girls From the Exploratory Structural Equation Modeling Analysis (Multiple-Group Gender—Model 5) at Wave 1

	Bully			Victim		
Factor	Verbal	Social	Physical	Verbal	Social	Physical
Bully–verbal	—	**.618**	**.237**	.299	.082	.082
Bully–social	**.433**	—	**.124**	.123	.145	.033
Bully–physical	**.580**	**.532**	—	.109	−.024	.383
Victim–verbal	.219	.083	.117	—	**.638**	**.317**
Victim–social	.051	.178	.019	**.535**	—	**.237**
Victim–physical	.169	.187	.315	**.601**	**.616**	—

Note. Correlations among the six bullying and victimizations factors above the main diagonal are for girls, while the correlations below the main diagonal are for boys. Correlations in bold are for three subdomains within the bullying domain and for the three subdomains within the victimization domain.

girls. For boys, correlations among all three bullying factors and among all three victimization factors are substantial (.433–.616), but the correlations between verbal and social factors are lower than between these two factors and the physical factors. For girls, the factors tend to be less highly correlated (.124–.638), but the correlations between the verbal and social factors (.618 for bullying, .638 for victimization) are much higher than correlations between these two factors and the physical factors. Girls who use physical bullying are much less likely to use social or verbal bullying, and there is a similar pattern of results for girls who are victims of bullying. Hence, consistent with some previous research (e.g., Card et al., 2008), the bully/victim factors are better differentiated for girls than for boys. Nevertheless, the pattern of correlations between the three bullying factors and the three victimization factors is similar across gender.

Latent factor mean comparison across gender. Finally, we are now in a position to address the issue of the invariance of the factor means across the two groups. In the taxonomy, the final four models (MGG–M10 through MGG–M13 in Table 2) all constrain mean differences between girls and boys to be zero—in combination with the invariance of other parameters. Again, there are several models that could be used to test gender mean invariance: MGG–M5 vs. MGG–M10; MGG–M7 vs. MGG–M11; MGG–M8 vs. MGG–M12; MGG–M9 vs. MGG–M13. However, from earlier models, the comparison based on MGG–M5 and MGG–M10 appears most critical (as there was no support for the invariance of factor uniquenesses or factor variances–covariances). Evaluation-of-fit indices for each of these pairs of model comparisons shows a consistent pattern suggesting that at least some of the six Adolescent Peer Relations Inventory (APRI) latent means do differ systematically for girls and boys. In the ESEM multiple group approach, latent means are typically constrained to be zero in one group and freely estimated in other groups. Inspection of group means based on MGG–M5 demonstrates substantial gender differences; for both the bully and victim domains, girls have much lower scores for the two physical (bully and victim) factors, moderately lower scores on the verbal factors, and almost no differences on the social scales. Hence, while boys have higher bully and victim scores than girls overall, there are substantial differences in types of bullying used by the two groups (also see subsequent discussion of gender differences in relation to Figure 1).

In summary, guided by the taxonomy of 13 factorial and measurement invariance models, the ESEM approach applied to the bullying and victimization factors based on responses to the APRI provides good support for the invariance of factor loadings and item intercepts (and thus strong factorial invariance, which is the absence of differential item functioning) but not for the invariance of uniquenesses, factor variances–covariances, and latent means across gender.

Invariance across year group. Does the APRI factor structure vary with year in school? Does the nature of bully and victim factors change during high school years? To address this issue, we evaluat the invariance of the APRI factor structure for Wave 1 responses over the five year-in-school groups, again applying the 13-model taxonomy of invariance. Because of the large number of groups, the degrees of freedom for the various comparisons are much larger than with the invariance over gender, but the logic is the same.

Inspection of the multigroup models associated with year in school (multiple-group year [MGY] models in Table 2) indicates good support for weak (MGY–M2 vs. MGY–M1) and strong (MGY–M5 vs. MGY–M2 and MGY–M1) measurement invariance. However, support for strict measurement invariance (MGY–M7 vs. MGY–M5) is lacking, suggesting that measurement error differs over year groups. However there is reasonable support for the invariance of the factor variance–covariance matrix (e.g., MGY–M4 vs. MGY–M2 or MGY–M8 *vs.* MGY–M5). Finally, the observed changes in fit for Models MGY–M5 to MGY–M10 in which latent means are constrained to be invariant over the year group all remain small and smaller than the recommended cutoff scores. However, given the number of means that are compared and the fact that fit still did slightly decrease when latent means are constrained to equality (for instance, the TLI decreases from .936 to .930 in Model MGY-5 vs. Model MGY–10), the results again suggest that at least some of the latent means may differ across year in school groups. Based on MGY–M5 (strong factorial invariance), we analyzed these latent mean differences, with latent means constrained to be zero in the first group (Year = 7). For all six factors, scores tend to be lowest in Year 7, increase in Year 8, remain reasonably stable in Years 9 and 10, and then decline in Year 11. However the pattern of results differs

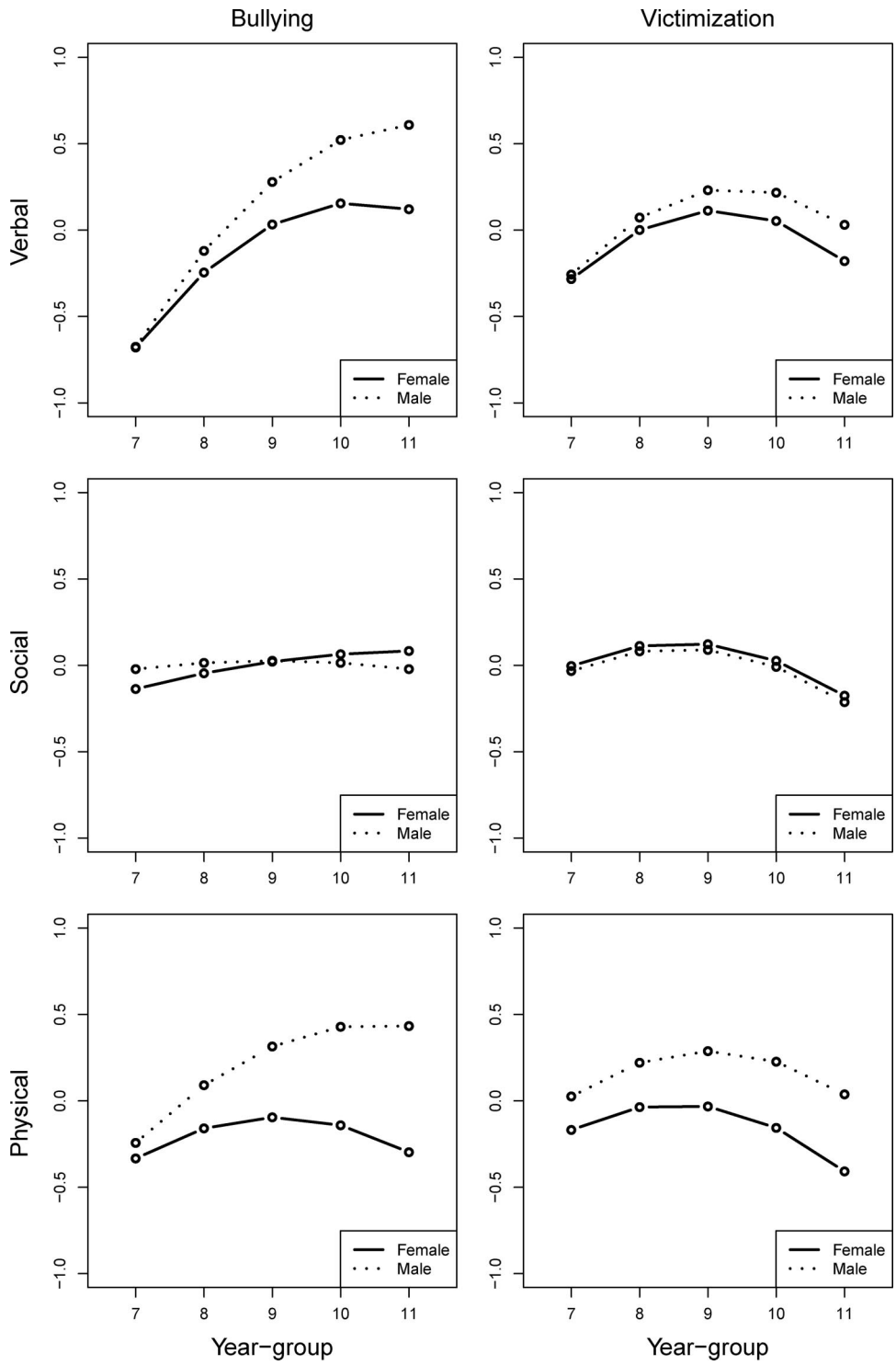

Figure 1. Multiple indicator–multiple cause model displaying the effects of gender (boys vs. girls) and year–group effects on six bully and victim factors based on an exploratory structural equation modeling approach.

somewhat for the different APRI factors (also see subsequent discussion in relation to Figure 1).

Construct validity of bully/victim factors: Relations with substantively important correlates. How similar or different are bully and victim factors in relation to the other psychosocial variables? In support of discriminant validity of the six bully/victim factors, is there differentiation among the (verbal, social, physical) subdomains as well as domains (bully vs. victim)? In order to pursue this important issue, we select a set of additional constructs hypothesized to be related to bully and victim factors. These constructs are also likely to be useful in evaluating interventions designed to reduce bullying.

An ESEM based on responses at Time 1 is used to evaluate the relations between the bully and victim factors, the multiple dimensions of self-concept, and other psychosocial constructs (see Appendix A). Hence, factors representing each construct inferred on the basis of multiple constructs are optimally weighted combinations of the items designed to measure the construct and corrected for unreliability. The ESEM is very large, positing 32 latent factors and four (single-indicator) demographic variables (representing gender and year in school) based on responses to a total of 172 items. Particularly given the size of the factor structure, the model is well defined in that every item loaded more substantially on the factor it is designed to measure (target loadings) than on other factors and the goodness of fit is reasonable (CFI = .923, TLI = .916, RMSEA = .019).

Demographics: Gender and year-group. The results from the regression of the various factors and gender, year and their interactions are reported in Table 5 and illustrated in Figure 1. These results show that boys tend to have higher scores on the bully and victim factors, as well as support for the discriminant validity of the subdomains in relation to gender. Relative to girls, boys have much higher scores for the physical (bully and victim) subdomains and somewhat higher scores for the verbal subdomains. However, for the social subdomain, there are no significant gender differences. These results are highly similar to those reported in the multiple group analysis of invariance over gender. The small, mostly nonsignificant Sex × Year-in-School interactions show that gender differences are consistent over year in school.

Year in school has significant linear and nonlinear relations with the bully/victim factors. For all six factors, scores tend to be lowest in Year 7, increase in Year 8, remain reasonably stable in Years 9 and 10, and then decline in Year 11. However, the pattern varies for the different bully/victim factors. The increases with year in school are stronger for the bully factors than for the victim factors and are stronger for the verbal factors than for the social or physical factors. To understand these year-in-school differences, it is important to recognize that the bully and victim factors are positively related: those with higher bully scores also tend to have higher victim scores, and those with higher victim scores tend to have higher bully scores. As students who are high on both bully and victim factors grow older, it is easier for bullies to target younger victims, and there are fewer older bullies to target them. Hence, bullying increases with year in school more than victimization (we return to these issues in Study 2 where we look at these relations over time).

Taken together, these results support the convergent and discriminant validity of both the domains (bully vs. victim) and the subdomains (verbal, social, physical) within each of these domains. Thus, for example, gender differences clearly support the discriminant validity of the subdomains (big differences in physical and, to a lesser extent, verbal subdomains, but none for social subdomain). However, the year-in-school effects support the need to differentiate both domain and subdomains. Increases with year in school are clearly larger for bully factors than victim factors, but these effects are largest for the verbal subdomain and very small for the social and physical subdomains.

Attitudes toward bullies and victims. Pro-bully and pro-victim attitudes are negatively correlated (−.40). Overall, pro-bullying attitudes increase with year in school, whereas pro-victim attitudes decrease. Although girls tend to hold more pro-victim attitudes than boys, these gender differences are similar across ages. Relations between bully and victim factors and these attitudes generally support our expectations but also offered a few surprises (Table 5). Not surprisingly, bully factors are strongly positively related to pro-bully attitudes and negatively related to pro-victim attitudes. More interesting, however, is the finding that the victim factors have a similar (although weaker) pattern of relations. Victim factors are weakly but positively related to pro-bully attitudes and are nearly unrelated to pro-victim attitudes. For both bully and victim factors, pro-victim attitudes are most negatively related to the physical factors. For both bully and victim factors, pro-bully attitudes are more negatively related to physical and verbal subdomains than to the social subdomain. The low pro-victim attitudes by students with high victim scores provide further support for the negative feelings that these students have for themselves. The positive relation between the victim factor and pro-bully attitudes is also consistent with the suggestion that victimization might lead to subsequent bullying.

Participant roles in bullying situations. Four participant role scales (Active Bully Reinforcer, Passive Bully Reinforcer, Ignore/Disregard, and Target Advocate; see Appendix C) are based on students being asked to report their reaction to witnessing a bullying situation. They were asked, for example, whether they would encourage the bully actively by joining in or passively by staying to watch or whether they would ignore the incident or actually help the victim. The results show that advocating for the victim decreases with year in school, whereas active and particularly passive reinforcement of the bully tends to increase. Girls are more likely to advocate for the victim, whereas boys are more likely to take the active or passive reinforcer role when confronted with a bully situation. However, the nature of these gender differences is consistent over year in school.

As expected, the bully factors are most positively related to the active and passive reinforcement roles, but negatively related to ignoring and particularly to advocating for the victim (Table 5). However, these effects vary for the subdomains of the bully factors. Thus, the active reinforcing role is most strongly related to the bully–physical factor, while the passive and victim-advocate role are more strongly related to the bully–verbal factor. Again, the pattern of results for the victim factors is somewhat surprising and disturbing. The victim factors (particularly in the physical subdomain) are positively correlated with actively reinforcing the bully but are essentially unrelated to passively reinforcing, ignoring, or advocating for the victim. Again, this suggests that victims might identify with bullies more than with other victims, maybe even aspiring to become bullies, or simply willing to reinforce the bullies in an effort to avoid becoming the next victim.

Table 5
Bully and Victim Factors: Relations With Other Constructs

Factors	Bully factors			Victim factors		
	1	2	3	4	5	6
Bully						
1. Verbal	—					
2. Social	.42	—				
3. Physical	.53	.44	—			
Victim						
4. Verbal	.27	.08	.15	—		
5. Social	.05	.17	.01	.43	—	
6. Physical	.22	.15	.38	.51	.52	—
Participation						
7. Active Reinforcement	.45	.48	.52	.09	.09	.21
8. Victim Advocate	−.33	−.15	−.19	.00	.08	−.06
9. Passive Reinforcement	.38	.26	.24	.06	−.01	.04
10. Ignore	−.10	−.05	−.15	−.03	.04	−.03
Coping styles						
11. Avoidance	.11	.17	.10	.25	.33	.23
12. Problem Solving	−.13	−.02	−.08	−.04	−.02	−.01
13. Social Support	−.15	.03	.22	−.03	.02	−.15
Attitudes						
14. Pro-Bully	.49	.39	.47	.08	.03	.16
15. Pro-Victim	−.27	−.20	−.34	.06	.04	−.07
Locus of control						
16. Internal	−.06	−.15	−.18	−.01	−.07	−.08
17. External	.16	.16	.08	.22	.26	.24
Anger management						
18. Control	−.19	−.12	−.20	−.11	−.05	−.09
19. Internalize	.02	.04	−.06	.32	.33	.19
20. Externalize	.30	.20	.29	.11	.09	.14
Depression						
21. Depression	.10	.14	.07	.38	.40	.26
Nonacademic self-concept						
22. Physical	−.07	−.01	.01	−.11	−.03	−.02
23. Appearance	−.01	.04	.10	−.13	−.08	.03
24. Opposite Sex	.12	.09	.12	−.13	−.12	−.09
25. Same Sex	−.07	−.03	−.04	−.35	−.39	−.32
26. Parent	−.22	−.14	−.11	−.10	−.11	−.07
27. Honesty	−.49	−.33	−.36	−.13	−.06	−.19
28. Emotional	.06	−.07	.10	−.19	−.24	−.10
29. Esteem	−.20	−.09	−.14	−.11	−.11	−.07
Academic self-concept						
30. Math	−.11	−.09	−.09	−.07	−.09	−.04
31. Verbal	−.15	−.10	−.21	−.06	−.06	−.06
32. Academic	−.21	−.17	−.23	−.12	−.13	−.13
Demographic variables[a]						
33. Sex	−.23	−.03	−.38	−.11	.03	−.30
34. School Year	.28	.05	.09	.08	−.02	.00
35. Year^2	−.14	−.02	−.07	−.13	−.09	−.09
36. Sex × Year	−.07	.03	−.09	−.02	.01	−.03

Note. The first 32 constructs are latent factors (based on multiple indicators, 168 items in total) from exploratory structural equation modeling. $r \geq .04$ is significant at $p < .05$.
[a] See also Figure 1.

Coping styles. In three coping style (avoidance, problem solving, or support seeking; see Appendix C) scales, students are asked how they cope with a problem or stressful situation. These measures do not assess coping with bullying per se—rather, students are asked how they cope with general difficulties. Both bully and (particularly) victim factors are more positively related to use avoidant coping strategies (e.g., "I avoid the problem by watching television") and less correlated with problem solving or seeking social support (Table 5). However, avoidance is more positively correlated with victim factors (particularly in the physical and social subdomains) than with the bully factors. Whereas the bullying factors are negatively correlated with the problem solving and seeking social support factors, the victim factors are less related to these coping styles.

Locus of control. Internal and external locus of control scales have a surprisingly similar pattern of relations to the bully and to the victim factors (Table 5). External locus of control is positively related to both bully and particularly victim factors, while internal

locus of control tends to be negatively related to both bully and victim factors. The pattern of high external and low internal scores is not surprising for the victim factors, but is not the expected pattern for the bully factors. Indeed, internal control is even more negatively correlated with bullying than with victimization, which suggests that bullying behaviors may represent attempts to regain control over environments perceived as uncontrollable. The results, however, do reinforce the suggestion that those high on bully factors are similar to those high on victim factors in many respects.

Anger management. Three anger management factors (control, internalize, externalize; see Appendix C) show that both bullies and victims have difficulties in controlling their anger, particularly in the verbal or physical subdomains. However, a clear pattern emerges (Table 5) in which bully factors (particularly in the verbal and physical subdomains) are more correlated with externalizing anger (e.g., "I let it all out") and victim factors (particularly in the verbal and social subdomains) are more correlated with internalizing anger (e.g., "No one can tell I am furious inside"). Indeed, bully factors are almost unrelated to internalizing anger. Thus, both the domains and all the subdomains are important in understanding relations with anger management.

Depression. A substantial body of literature attests to the negative consequences of victimization in terms of depression (Hawker & Boulton 2000) but also strong associations between bullying and depression (e.g., Angold, Costello, & Erkanli, 1999; Kovacs & Devlin, 1998). Indeed, our results (based on the 10-item Childhood Depression Inventory) indicate that depression is positively related to particularly the victim factors, but also the bully factors (Table 5). In both cases, depression is more negatively related to the social and, to a lesser extent, verbal subdomains than to the physical subdomain. While bullying is often seen in physical terms, it is apparently the social and verbal domains that are most psychologically harmful. Hence, in understanding relations with depression, it is important to differentiate among subdomains as well as domains.

Multiple dimensions of self-concept. Eleven self-concept scales were considered—eight nonacademic factors (Physical Abilities, Appearance, Opposite Sex Relations, Same Sex Relations, Parent Relations, Honesty/Trustworthiness, Emotional Stability, and General Self-Esteem) and three academic factors (Math, Verbal, School; see Appendix A). Consistent with previous research, victim factors are consistently and negatively correlated with multiple dimensions of self-concept (Table 5). However, previous research did not provide clear expectations for the bullying factors. Although most of the observed correlations are negative, they are close to zero for the two physical self-concept factors, for the two peer relationships factors, and for the emotional stability factor. Interestingly, however, global self-esteem is negatively correlated with both the bully and victim factors and the sizes of these negative correlations are very similar. Hence, neither bullies nor victims seem to have particularly good self-concepts.

Despite the generally negative correlations of self-concept with both the bully and victim factors, there are some clear distinctions in the patterns of relations. Particularly notable and consistent with the Marsh et al. (2001) study of U.S. students, the bully factors are positively correlated with opposite sex relationships, meaning that bullies perceive themselves to be popular with the opposite sex. The victim factors are most negatively correlated with same sex relationship self-concepts—more so than other areas of self-concept in relation to either bully or victim factors. Emotional stability is also more negatively correlated with victim factors than with bully factors, which is consistent with finding on the depression scores discussed earlier. Honesty/trustworthiness self-concept is most negatively related to the bully factors. This suggests, perhaps, that bullies are cognizant that they are not doing the right thing when they bully other people, although other interpretations of this relation might also be viable. More generally, bully factors have moderate to strong negative correlations with honesty/trustworthiness, parent relationships, verbal, math, and school self-concepts, and these correlations are more negative than those observed for the victim factors. In summary, both bully and victim factors tend to be negatively related to the set of self-concept factors.

There are also interesting patterns in relation to subdomains. Particularly for the nonacademic components of self-concept and for self-esteem, the most negative relations tend to occur for the social and particularly the verbal subdomains of both the bully and victim domains, but not for the physical subdomain. Thus, for example, victim factors are most negatively correlated with the same-sex self-concept factor, but the correlations are more negative for verbal and particularly the social domains than for the physical domain. The next most negative correlations with the victim factors are with emotional self-concept, but here again the correlations are more negative for the verbal and particularly for the social subdomain than for the physical subdomain. Honesty is most negatively correlated with bully factors, but here it is for the verbal subdomain that the negative correlation is strongest.

In summary, although there are qualitative differences between relations with the bully and victim factors, the results suggest that there are many similarities between bullies and victims. Taken together, the pattern of results supports the need to consider both the domain and subdomains in order to understand relations between multiple dimensions of self-concept and the bully/victim factors.

Study 2

Study 2 was based on three waves of data collected during a single school year (Wave 1 of this data was the basis of Study 1). Our purposes in Study 2 were to evaluate the longitudinal stability of the a priori factor structure of the APRI measures of bully/victim factors and support for their convergent and discriminant validity in relation to time. We begin by evaluating the longitudinal factor structure, with a particular focus on the invariance of the factor structure over time and the stability of the three bully and three victim factors. We then apply a multitrait–multimethod (MTMM) design with time as the method factor to critically evaluate support for the convergent and discriminant validity of these six factors. Consistent with our substantive-methodological focus, we demonstrate substantively important implications of the results and the superiority of the new ESEM approach compared with traditional CFA approaches. Finally, we apply an autoregressive cross lagged panel analysis (Marsh & Craven, 2006; Little, Preacher, Selig, & Card, 2007) in an attempt to evaluate the directionality of the associations between bullying and victimization over time: Does prior bullying lead to subsequent victimization? Does prior victimization lead to subsequent bullying? Are bullying and victimization reciprocally related?

Method

Because Study 2 was based on longitudinal data from the same students as Study 1, the reader is referred to the Method section in Study 1 for a description of the participants, the measures, and statistical issues (e.g., the ESEM approach and goodness of fit) common to the two studies. Here we focus on the longitudinal statistical analyses that are specific to Study 2.

Invariance over time: Latent mean structure approach. In the evaluation of stability over time, researchers typically focus on either covariance stability (e.g., test–retest correlations) or mean stability (changes in the level of a construct over time). Historically, both approaches to stability were limited due to reliance on manifest measures that confounded measurement error and typically untested assumptions of measurement invariance over time (Marsh, 1993b; Marsh & Grayson, 1994; Meredith, 1993). However, with minor adaptations, it is possible to apply the same set of 13 models in the taxonomy presented in Study 1 (Table 1) to test the invariance of the six-APRI-factor structure over time using the ESEM approach to longitudinal data (Marsh et al., 2009; Marsh, Lüdtke, Muthén et al., 2010). Within this framework, given underlying assumptions, it is possible to test both covariance stability in relation to latent constructs and the stability of latent mean differences in the same model. In both approaches, a critical test is whether differences in item intercepts can be explained in terms of differences in latent means. Unless mean differences are reasonably consistent across the items used to infer a construct (as reflected in support for the invariance of intercepts), then mean differences at the latent construct level are likely to be an idiosyncratic function of the choice of the particular items chosen to represent a latent construct. This problem is called *differential item functioning* as the items function differently in terms of latent mean differences. For comparisons between gender groups, for example, this would mean that gender differences are not consistent across items used to measure a construct. In the case of longitudinal data, the comparison of means over time becomes problematic in that the meaning of the items has apparently shifted over time (Marsh & Grayson, 1994). In the present investigation, we explored whether there is evidence for differential item functioning for the 36 APRI items. Because of this focus on differential item functioning, Model 5 in the taxonomy (see Table 1) is critical and is a particular focus of Study 2.

In this ESEM application with longitudinal data over three waves, we posited sets of 36 cross-wave correlated uniquenesses to account for the residual associations between matching items administered at Time 1 (T1), T2, and T3. Indeed, when the same item is used on multiple occasions, a correlation is likely to exist between the unique components of each item on the two occasions that cannot be explained by the correlations between the factors. The failure to include these a priori correlated uniquenesses (CUs) is likely to systematically bias parameter estimates such that test–retest correlations among matching latent factors are systematically inflated (see e.g., Jöreskog, 1979; Marsh, 2007a; Marsh & Hau, 1996).

Multitrait–multimethod (MTMM) analysis of convergent and discriminant validity. The MTMM design is used widely to assess convergent and discriminant validity, and is one of the standard criteria for evaluating psychological instruments (e.g., Byrne, 1996; Marsh & Hattie, 1996; Shavelson, Hubner & Stanton, 1976; Wylie, 1989). The MTMM design provides a particularly strong approach to evaluating stability of responses to a multidimensional instrument, as emphasized by Campbell and O'Connell (1967) who specifically operationalized the multiple methods in their MTMM paradigm as multiple occasions. Marsh (Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Marsh, Martin, & Jackson, 2010) also recommended this approach to evaluate support for the convergent and discriminant validity in relation to temporal stability over time.

Campbell and Fiske's (1959) paradigm is, perhaps, the most widely used construct validation design. Although their original guidelines are still widely used to evaluate MTMM data, important problems with their guidelines are well known (see reviews by Marsh, 1988, 1993a; Marsh & Grayson, 1995). Ironically, even in highly sophisticated CFA approaches to MTMM data, a single scale score—often an average of multiple items—is typically used to represent each trait–method combination. Marsh (1993a; Marsh et al., 2005; Marsh & Hocevar, 1988), however, argued that it is stronger to incorporate the multiple indicators explicitly into the MTMM design. When multiple indicators are used to represent each scale, CFAs at the item level results in a MTMM matrix of latent correlations, thereby eliminating many of the objections to the Campbell–Fiske guidelines. In the present investigation, we apply this approach with both CFA and ESEM, providing a strong test of our claim that ESEM provides systematically stronger support for discriminant validity of the latent factors than does CFA.

Results

Invariance over time: Latent mean structure approach. We began by testing the configural invariance of responses with and without the sets of correlated uniquenesses between matching items at T1, T2 and T3 (Models Multigroup Time [MGT]–M0 and MGT–MI in Table 2). The results clearly demonstrate that the inclusion of the 108 correlated uniquenesses (3 waves × 36 items) improves the fit substantially (.958 vs. .936 for TLI; .962 vs. .941 for CFI; .016 vs. .019 for RMSEA). Consistent with a priori expectations, the test–retest correlations of the APRI factors in MGT–M0 are positively biased, although the differences are not large; the mean test–retest correlation for adjacent occasions is .629 for MGT–M0 and .603 for MGT–M1. Based on this initial analysis in support of configural invariance with CUs, we subsequently consider models that include CUs for the 36 matching APRI items for T1, T2, and T3. We now focus on the key 13 tests of measurement invariance (Models MGT–M1 through MGT–M13 in Table 2) to the APRI factor structure over time.

Inspection of the multiwave models (MGT models in Table 2) indicates good support for weak (MGT–M2 vs. MGT–M1), strong (MGT–M5 vs. MGT–M2 and MGT–M1), and strict (MGT–M7 vs. MGT–M5) measurement invariance. In addition, there is reasonable support for the invariance of the factor variance–covariance matrix (e.g., MGT–M4 vs. MGT–M2 or MGT–M9 *vs.* MGT–M7) and of the latent means (see MGT–M10 to MGT–M13). These findings suggest that the six APRI factor means do not differ substantially over time for the three waves of data collected in a single school year considered here.

MTMM tests of convergent and discriminant validity. Now we turn to a unique application of the ESEM model to test the discriminant validity of the ESEM factors in relation to time and to compare it to the discriminant validity of the corresponding CFA

factors. We begin with a fully latent MTMM matrix based on the strong invariance model (Model MGT–M5). An important feature of this study is the use of the MTMM approach to systematically compare results based on the ESEM model already considered with that of the corresponding CFA model. Consistent with results from the ESEM models, the fit of the CFA model positing strong factorial invariance over time (factor loadings and intercepts invariant) is reasonable (CFI =.937; TLI=.935; RMSEA = .020) but not as good as the corresponding ESEM model (CFI = .961; TLI = .959; RMSEA = .016). The critical issue here is to test our a priori prediction that support for the discriminant validity for the CFA model is substantially poorer than for the ESEM model.

The application of the MTMM logic here is somewhat more complex than in the typical MTMM design that has only two facets (traits and methods). However, here we have a three-facet design: trait domain (bully and victim); trait subdomain (verbal, social, and physical); method (represented by the three occasions). Thus, we actually have two distinct "instruments" (bully and victim), each containing verbal, social, and physical subdomains. However, for example, the physical bullying factor is clearly a different construct than the physical victimization factor. Also, while the three time waves are considered as method factors, it is reasonable to assume that relations between matching factors should be higher for adjacent time points (e.g., physical bullying at T1 and T2 or at T2 and T3) than at nonadjacent time points (e.g., physical bullying at T1 and T3). Consistent with these various distinctions, we classified correlations into 11 categories (rather than the four used by Campbell and Fiske, 1959) based on various combinations of traits, time (near or far comparisons), and instruments (same or different) as depicted in Table 6.

For both the CFA and ESEM results, there is good support for convergent validity (stability over time), particularly for adjacent time points (Time = near in Table 7). Although support for convergent validity is slightly higher for the CFA model than the corresponding ESEM model (.624 vs. .616 for near convergent validities in Type a; .490 vs. .481 for far convergent validities in Type b), these differences are very small.

In the most demanding test of discriminant validity in the MTMM approach, convergent validities are compared with correlations among different traits measured with the same method (or same occasion and same instrument here). While there is reasonable support for the discriminant validity of ESEM factors (mean convergent validity in Type a = .616, compared with mean comparison correlations in Type c = .525), there is a complete lack of support for the discriminant validity based on the CFA factors (mean convergent validity in Type a = .624, compared with mean comparison correlation in Type c = .818).

Support for discriminant validity of the APRI factors varies depending upon whether comparison correlations are based on different factors collected at the same time (Type c), adjacent near times (Type d), or nonadjacent "far" times (Type e), and whether convergent validities are based on convergence of same traits administered in adjacent near times or nonadjacent far times. However, across all combinations of these convergent validities

Table 6
Summary of Multi-Trait Multi-Method (MTMM) Matrix of Correlations: ESEM Above the Main Diagonal, CFA Below the Main Diagonal Based on Model 5 (Type of Correlation in Parentheses)

	Time 1						Time 2						Time 3					
	Bully			Victim			Bully			Victim			Bully			Victim		
Time/trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Time 1																		
1. BV	—	.47^{c}	.53^{c}	.27^{f}	.05^{i}	.21^{i}	**.67^{a}**	.34^{d}	.41^{d}	.20^{g}	.06^{j}	.20^{j}	**.57^{b}**	.26^{e}	.35^{e}	.19^{h}	.07^{k}	.15^{k}
2. BS	.73^{c}	—	.48^{c}	.10^{i}	.16^{f}	.18^{i}	.32^{d}	**.54^{a}**	.32^{d}	.08^{j}	.17^{g}	.16^{j}	.28^{e}	**.35^{b}**	.25^{e}	.10^{k}	.13^{h}	.14^{k}
3. BP	.83^{c}	.73^{c}	—	.16^{i}	.01^{i}	.39^{f}	.36^{d}	.29^{d}	**.67^{a}**	.11^{j}	.02^{j}	.35^{g}	.30^{e}	.24^{e}	**.56^{b}**	.14^{k}	.10^{k}	.33^{h}
4. VV	.26^{f}	.16^{i}	.21^{i}	—	.53^{c}	.53^{c}	.20^{g}	.09^{j}	.15^{j}	**.62^{a}**	.33^{d}	.40^{d}	.16^{h}	.07^{k}	.13^{k}	**.52^{b}**	.26^{e}	.29^{e}
5. VS	.15^{i}	.18^{f}	.14^{i}	.84^{c}	—	.40^{c}	−.03^{j}	.14^{g}	.04^{j}	.35^{d}	**.54^{a}**	.38^{d}	−.02^{k}	.08^{h}	.03^{k}	.30^{e}	**.38^{b}**	.27^{e}
6. VP	.27^{i}	.22^{i}	.34^{f}	.81^{c}	.83^{c}	—	.15^{j}	.17^{j}	.38^{g}	.38^{d}	.39^{d}	**.68^{a}**	.16^{k}	.16^{k}	.33^{h}	.33^{e}	.31^{e}	**.52^{b}**
Time 2																		
7. BV	**.67^{a}**	.48^{d}	.57^{d}	.20^{g}	.10^{j}	.22^{j}	—	.40^{c}	.51^{c}	.28^{f}	.00^{i}	.22^{i}	**.65^{a}**	.29^{d}	.38^{d}	.24^{g}	.04^{j}	.16^{j}
8. BS	.46^{d}	**.57^{a}**	.44^{d}	.14^{j}	.15^{g}	.19^{j}	.74^{c}	—	.56^{c}	.12^{i}	.28^{f}	.28^{i}	.28^{d}	**.49^{a}**	.26^{d}	.13^{j}	.23^{g}	.23^{j}
9. BP	.57^{d}	.47^{d}	**.67^{a}**	.18^{j}	.13^{j}	.31^{g}	.86^{c}	.77^{c}	—	.16^{i}	.12^{i}	.51^{f}	.37^{d}	.33^{d}	**.66^{a}**	.18^{j}	.16^{j}	.41^{g}
10. VV	.21^{g}	.14^{j}	.17^{j}	**.64^{a}**	.54^{d}	.56^{d}	.29^{f}	.23^{i}	.26^{i}	—	.53^{c}	.56^{c}	.18^{g}	.05^{j}	.12^{j}	**.65^{a}**	.35^{d}	.36^{d}
11. VS	.13^{j}	.18^{g}	.13^{j}	.52^{d}	**.59^{a}**	.55^{d}	.19^{i}	.29^{f}	.24^{i}	.85^{c}	—	.57^{c}	.00^{j}	.15^{g}	.05^{j}	.34^{d}	**.52^{a}**	.37^{d}
12. VP	.24^{j}	.19^{j}	.30^{g}	.56^{d}	.56^{d}	**.68^{a}**	.33^{i}	.31^{i}	.44^{f}	.84^{c}	.84^{c}	—	.17^{j}	.21^{j}	.37^{g}	.41^{d}	.39^{d}	**.62^{a}**
Time 3																		
13. BV	**.55^{b}**	.39^{e}	.47^{e}	.16^{h}	.09^{k}	.20^{k}	**.65^{a}**	.45^{d}	.56^{d}	.19^{g}	.12^{j}	.25^{j}	—	.38^{c}	.48^{c}	.27^{f}	.04^{i}	.19^{i}
14. BS	.34^{e}	**.38^{b}**	.34^{e}	.10^{k}	.11^{h}	.16^{k}	.43^{d}	**.51^{a}**	.44^{d}	.12^{j}	.16^{g}	.21^{j}	.78^{c}	—	.64^{c}	.18^{i}	.42^{f}	.40^{i}
15. BP	.46^{e}	.36^{e}	**.54^{b}**	.15^{k}	.11^{k}	.26^{h}	.54^{d}	.40^{d}	**.62^{a}**	.17^{j}	.14^{j}	.31^{g}	.87^{c}	.82^{c}	—	.22^{i}	.25^{i}	.56^{f}
16. VV	.20^{h}	.14^{k}	.20^{k}	**.52^{b}**	.43^{e}	.46^{e}	.27^{g}	.21^{j}	.26^{j}	**.64^{a}**	.52^{d}	.56^{d}	.35^{f}	.33^{i}	.36^{i}	—	.50^{c}	.57^{c}
17. VS	.14^{k}	.15^{h}	.18^{k}	.40^{e}	**.45^{b}**	.42^{e}	.19^{j}	.25^{g}	.25^{j}	.53^{d}	**.58^{a}**	.56^{d}	.30^{j}	.43^{f}	.38^{i}	.86^{c}	—	.66^{c}
18. VP	.21^{k}	.17^{k}	.30^{h}	.41^{e}	.41^{e}	**.51^{b}**	.28^{j}	.26^{j}	.38^{g}	.53^{d}	.51^{d}	**.63^{a}**	.39^{i}	.43^{i}	.52^{f}	.85^{c}	.89^{c}	—

Note. The correlations in bold are convergent validities (correlations between the same trait administered on different occasions when the method is time). Consistent with the design of the analysis, correlations are classified into 11 types represented by the superscripts a–k (see Table 7 for a description). ESEM = exploratory structural equation modeling; CFA = confirmatory factor analysis; BV = bully–verbal; BS = bully–social; BP = bully–physical; VV = victim–verbal; VS = victim–social; VP = victim–physical.

Table 7
Summary of Multitrait Multimethod Correlations

Type	Trait	Time	Instrument	No. of correlations	MGT–M5 ESEM	MGT–M5 CFA	MGT-M5 ESEM	MGT-M5 CFA
			Convergent validities					
A	Same	Near	Same (MtHmMi–near)	12	.616	.624	.602	.619
B	Same	Far	Same (MtHmMi–far)	6	.481	.490	.455	.470
Total				18	.568	.578	.553	.569
			Comparison coefficients					
C	Different	Same	Same (HtMmMi)	18	.525	.818	.516	.819
			Comparison coefficients					
D	Different	Near	Same (HtHmMi–near)	24	.351	.513	.331	.511
E	Different	Far	Same (HtHmMi–far)	36	.288	.409	.255	.390
			Matching bully and victim traits					
F	Match	Same	Different (HtMmHi)	9	.351	.345	.360	.350
G	Match	Near	Different (HtHmHi–near)	12	.328	.321	.327	.321
H	Match	Far	Different (HtHmHi–far)	6	.202	.199	.184	.187
			Nonmatching bully and victim traits					
I	Different	Same	Different (HtMmHi)	18	.161	.271	.171	.277
J	Different	Near	Different(HtHmHi–near)	27	.133	.198	.133	.200
K	Different	Far	Different (HtHmHi–far)	12	.103	.157	.090	.145
Total				153	.299	.397	.291	.394

Note. Each correlation from the extended multitrait multimethod matrix was classified into 11 types depending the trait, method (time), and instrument. Traits were the same (same trait, same instrument), different, or matching (matching traits from different instruments (e.g., bully–physical vs. victim–physical). Time was the same, near (T1 vs. T2 or T2 vs. T3) or far (T1 vs. T3). Instrument was the same (both bully or both victim) or different (bully vs. victim). MtHmMi–near = mono-trait hetero-method mono-instrument–near time; MtHmMi–far = mono-trait hetero-method mono-instrument–far time; HtMmMi) = hetero-trait mono-method mono-instrument; HtHmMi = hetero-trait hetero-method mono-instrument; HtMmHi = hetero-trait mono-method hetero-instrument; HtHmHi = hetero-trait hetero-method hetero-instrument.

and comparison correlations, the support for the discriminant validity is systematically stronger for ESEM factors than CFA factors (Table 7). In summary, these results support the convergent and discriminant validity of the ESEM factors and provide clear evidence for our claim that ESEM factors are substantially better than the corresponding CFA factors in terms of discriminant validity.

Directional ordering of the bully and victim factors over time. Does bullying lead to subsequently victimization, or vice versa, or are the two reciprocally related such that each leads to the other? Questions like these have important practical implications to the evaluation of the directional ordering of the bully and victim factors over time. There is a well-established paradigm for evaluating the directionality of the associations when the same constructs are measured on multiple occasions based on SEMs (e.g., Little et al., 2007; Marsh & Craven, 2006), but here we illustrate how these questions can be addressed with ESEM. As in the preceding section, we begin the Model MGT–M5, in which we estimated correlations among the 18 APRI factors (six factors from three occasions). In the present application, we reparameterize the model and represent all of the longitudinal correlations with path coefficients. Thus, for example, the $6 \times 6 = 36$ correlations between T1 factors and T2 factors are represented as 36 path coefficients. Correlations among factors in the same wave are still estimated as factor correlations (or correlations between factor residuals). In this respect, this "full-forward" autoregressive cross-lagged ESEM model is equivalent to the corresponding measurement model in that the goodness of fit, degrees of freedom, and number of estimated parameters are equal between the two models.

Because we have already evaluated the factor structure of this model in detail, we focus on path coefficients (Table 8). As the same variables were administered on three occasions, we expected that much of the effect of T1 measures on T3 measures would be mediated by T2 measures. In order to examine the implications of this distinction, we obtained separate estimates of direct effects (the typical path coefficient), indirect effects mediated through intervening variables, and total effects that are merely the sum of direct and indirect effects (see related discussion by Little et al., 2007; Marsh & O'Mara, 2010). We expected that the effects of T1 factors on T3 factors would be substantially mediated by T2 factors.

Inspection of results shows, not surprisingly, that the strongest paths are between the same factors on adjacent occasions (the values in bold in Table 8). Although similar to stability correlations considered earlier (Table 7), these tend to be smaller because the model controls for the effects of other variables. Particularly at T3, the majority of the effects from T1 are mediated by T2 factors. There are, however, a number of additional paths (direct and indirect effects) that are statistically significant.

T1 physical bullying has a significant positive effect on physical victimization at T2 and T3 (although the effect at T3 is mostly mediated through T2 scores). Similarly, T1 physical victimization

Table 8
Direct (D) and Total (T) Effects Leading From Times 1 and 2 to Times 2 and 3: Parameter Estimate and Standard Errors (Based on Model 5)

	From Time 1												From Time 2											
	Bully						Victim						Bully						Victim					
	BV		BS		BP		VV		VS		VP		BV		BS		BP		VV		VS		VP	
Variable	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*
To Time 2																								
BV D	**.65**	**.04**	.02	.04	−.02	.05	.06	.03**	−.14	.04	.06	.04												
BS D	.11	.05	**.48**	.05	−.01	.06	−.05	.03	.05	.04	.07	.05												
BP D	.10	.04	−.01	.05	**.56**	.05	−.04	.03	−.05	.04	.19	.04												
VV D	.05	.04	.00	.03	−.04	.04	**.57**	.04	−.01	.04	.08	.04**												
VS D	−.01	.03	.12	.03	−.11	.04	.03	.04	**.41**	.05	.19	.05												
VP D	.01	.04	−.02	.04	.11	.04	.04	.04	.05	.05	**.60**	.05												
To Time 3																								
BV D	**.22**	.06	.05	.06	−.07	.07	−.02	.04	−.04	.04	.09	.05*	**.49**	.06	−.01	.05	.05	.07	.01	.04	−.03	.04	−.02	.06
BV T	**.54**	.05	.05	.04	−.06	.06	.02	.03	−.12	.04	.11	.05												
BS D	.01	.06	**.07**	.05	.05	.06	.02	.04	−.01	.04	.02	.05	.16	.06*	**.40**	.06	−.06	.08	−.11	.04	.03	.04	.10	.06
BS T	.12	.05	**.26**	.05	.03	.05	−.05	.03	.01	.04	.10	.04												
BP D	.01	.06	.06	.05	**.14**	.07*	−.01	.04	.01	.04	.04	.05	.08	.06	−.18	.05	**.57**	.08	−.04	.04	−.01	.04	.05	.06
BP T	.10	.04	−.04	.04	**.47**	.05	−.04	.03	−.04	.04	.17	.04												
VV D	−.03	.04	.03	.04	.01	.05	**.17**	.05	.03	.04	.02	.05	.05	.04	−.01	.04	.05	.06	**.53**	.05	−.01	.05	−.01	.06
VV T	.03	.04	.02	.03	.00	.04	**.47**	.04	.01	.04	.07	.04												
VS D	.00	.06	.05	.04	.06	.06	−.02	.05	**.14**	.06	.00	.06	−.06	.06	.10	.05	.02	.07	.10	.05**	**.34**	.05	.05	.07
VS T	−.02	.04	.04	.04	.04	.04	.04	.04	**.30**	.05	.11	.05												
VP D	−.07	.06	−.03	.05	.11	.06*	.01	.05	.00	.05	**.14**	.02	.00	.06	.03	.05	.10	.07	.04	.05	.06	.04	**.39**	.07
VP T	−.05	.04	−.02	−.04	.20	.05	.04	.04	.04	.05	**.41**	.05												

Note. The results are based relations between six bully and victim factors collated on three occasions (also see Table 6). Direct effects are standardized. The coefficients in bold are for the effects of the same factor collected on two different occasion (also see Table 6 for the corresponding test–retest stability correlations). BV = bully–verbal; BS = bully–social; BP = bully–physical; VV = victim–verbal; VS = victim–social; VP = victim–physical; Est = estimate; *SE* = standard error; D = direct effects; T = total effects.

has significant effects on physical bullying at T2 and T3. This suggests a pattern of reciprocal effects such that being a victim of physical bullying leads to becoming a physical bully, while those who engage in physical bullying subsequently become victims of physical bullying themselves. However, interpretations of this pattern of results need to be tempered by the finding that there are no statistically significant paths from any of the T2 physical bullying or victimization factors to the T3 physical bullying or victimization factors. Thus, the substantial correlations between T2 and T3 measures are largely explained in terms of T1 measures, suggesting that T1 measures collected at the beginning of the school year may have "set the stage" for the remainder of the school year.

For the verbal and social factors, there are a few significant paths leading from bullying to victimization but not any leading from victimization to bullying. This suggests that the pattern of reciprocal effects might be specific to the physical subdomain. There are also several significant paths leading from verbal bullying to other bullying factors, but none leading from social or physical bullying to other bullying factors. However, there are no statistically significant positive paths leading from any of the victimization factors to the other victimization factors.

In summary, as already shown, the bully and victim factors are reasonably stable over time. While T1 factors are substantially correlated with corresponding factors at T2 and T3, most of the effects of T1 factors on T3 factors are mediated through T2 factors. Particularly for the physical domain, we found evidence of a pattern of reciprocal effects such that prior bullying leads to subsequent victimization even after controlling for prior victimization and that prior victimization leads to subsequent bullying even after controlling for prior bullying. Supplemental (unreported) multigroup analyses also demonstrate that these results generalize over responses by boys and girls.

Discussion

The present investigation is a substantive-methodological synergy (Marsh & Hau, 2007), bringing to bear new, strong, and evolving methodology to evaluate a new multidimensional instrument to measure bullying and victimization and to tackle complex substantive issues in this area of research. Here, as in other psychological research, good theory, measurement, research, and practice are inexorably related such that the neglect of one will undermine pursuit of the others. A particularly important starting point is the development of a solid measure of bullying and victimization, based on sound theory and empirical support for construct validity. Important contributions of the present investigation include the presentation of a psychometrically sound measure of multiple bully and victim factors, new approaches and evidence in support of the convergent and discriminant validity of this measure, and the illustration of the flexibility and potential contribution of new statistical methodology (ESEM) to educa-

tional psychology research. In support of our claim for the importance of substantive-methodological synergies, the application of a sound measure and new methodological approaches resulted in substantively important findings with practical implications for understanding the nature of bullying and victimization.

In discussion of our results and their implications for practice, we emphasize the importance of (a) convergent and discriminant validity in testing hypotheses about multidimensional constructs, (b) analysis of measurement invariance over multiple groups and over time, (c) detection of differential item functioning, and (d) the study of the stability of constructs over time. It is our contention that these important topics have not been given sufficient attention in applied quantitative research that appears in the leading substantive journals in educational psychology (but see Marsh & Hau, 2007). Based on the combination of attention to these important measurement principles and the ESEM approach to factor analysis highlighted here, we provide an important example of construct validation for educational psychology. Additionally, we highlight important issues to consider when developing measurement tools relevant to complex issues with important implications for theory and practice. Measurement, theory, research, and practice are inexorably intertwined such that the neglect of one will undermine the others. In developing the logic of this argument, we began with the development of a psychometrically sound measure, evaluated the construct (convergent and discriminant) validity in relation to other constructs and the application of MTMM analyses, and finally evaluated the directional ordering of bullying and victimization. At each stage, we demonstrated the importance of ESEM in pursuit of these aims and, thus, the importance of substantive-methodological synergy.

The development of new psychometrically sound instrumentation. Fundamental to the present investigation was the development of a new, psychometrically sound measure of bullying and victimization—coupled with the application of ESEM. On the basis of previous theory and research, we posited that bullying and victimization domains could both be further divided into three subdomains: verbal, social–relational, and physical. The APRI was specifically designed to measure these six a priori factors, and clear support for them was found. While there was reasonable support for the a priori factor structure based on a traditional ICM–CFA approach, the correlations among the bully factors and among the victim factors were so high as to seriously undermine ability to distinguish among the three subdomains in relation to both bullying and victimization, a result that appears to be the norm in this area of research (Card et al., 2008). In this respect, the application of ESEM made an important contribution, demonstrating that the factor correlations were much lower than those estimated with traditional approaches, such as ICM–CFA. While previous research provided some support for the discriminant validity for the physical, verbal, and social subdomains for aggressive behavior in general, the correlations among these subdomains were so high as to obscure relations with other variables. Furthermore, Card and Hodges (2008) reported that there were no previous measures of victimization that were psychometrically sound.

ESEM results also indicated that there were also modest positive correlations between bullying and victimization. This result is consistent with previous research and underpins the finding that bullying and victimization co-occur in the same person, as is demonstrated by positive relations between these latent variables. Here, however, we showed that these positive correlations were reasonably subdomain specific (e.g., physical bullying was more highly correlated with physical victimization than with other forms of victimization).

Construct validity: Relations with other constructs. There was a clear pattern of gender differences in the bully and victim factors (Figure 1) that was consistent with a priori predictions. Boys had higher scores on both the bully and victim factors, but the sizes of these differences varied substantially with the particular subdomain. The largest gender differences were for the physical factors, followed by the verbal factors. However, there were no gender differences in the social factors (e.g., exclusion, rumor spreading, and manipulating friendships). The gender differences were essentially parallel for the verbal, social, and physical subdomains of the bullying and victimization factors. These gender differences did not vary much as a function of year in school. Again, the contribution of ESEM was important in that the gender differences were larger and provided stronger support for a priori predictions based on ESEM than ICM–CFA factors. For example, the ICM–CFA factors showed that boys had higher scores than girls on the social factors and that gender differences in favor of boys were smaller for the physical and verbal subdomains. These results are consistent with our claim that ESEM factors are better differentiated (less correlated), providing better discriminant validity in relation to other variables. This pattern of results for bullying seems to be consistent with meta-analyses of aggressive behaviors. However, support for a largely parallel set of results for victimization is an apparently new finding with potentially important implications to theory and practice—particularly in relation to the dearth of psychometrically strong measures of victimization.

Our results also indicate that for all three bullying subscales, there were statistically significant linear and nonlinear effects of year in school (Figure 1). Bullying behaviors increased during the early high school years and then tended to level out in later high school years. Importantly, however, there was little decrease in bullying over these high school years. For the three victim factors, the linear effects were small, whereas the quadratic effects were substantial. In particular, there was a substantial increase for all three scales from Year 7 to Year 8, followed by a gradual decline in victimization in Years 10 and 11 (see Figure 1). These year-in-school differences did not vary much as a function of gender. As suggested by Smith et al. (1999), bullies may be more likely to target younger students, and with increasing years in school, some of the victims may learn strategies to counter bullying. These year-in-school differences did not vary much as a function of gender. There were also systematic differences in the different subdomains. Particularly verbal bullying showed substantial increases with year in school, while social bullying and social victimization were relatively flat over the year-in-school range. Physical bullying had the largest interaction between gender and year-in-school group, showing increases with year in school for boys but a reasonably flat quadratic (inverted U) function for girls. In summary, these results show that the forms that bullying and victimization take evolve with year in school but that, at least for the physical domain, these year-in-school effects also vary with gender. These differentiated patterns clearly support the need to measure separation domains and subdomains of bullying and victimization.

Importantly, through the application of latent-variable models based on ESEM, we were able to demonstrate support for strong measurement invariance that is an implicit—but typically untested—assumption of all studies that evaluate gender and year in school effects in bullying and victimization. Both in terms of the invariance of factor loadings and differential item functioning (intercept invariance), there was good support for the invariance of the APRI factor structure over gender, year in school, and time. However, particularly for gender, there was not good support for the invariance of factor correlations. Indeed, the pattern of correlations was quite distinct. For boys, correlations among the three bully factors and among the three victim factors were all substantial—suggesting, perhaps, the presence of higher order bully and victim factors. For girls, however, the physical factors were quite distinct from the social and verbal subdomains for both bullying and victimization. Nevertheless, the pattern of correlations relating the three bully factors to the three victim factors was similar for boys and girls, as well as their pattern of associations over time.

In support of the construct validity of this measure of bullying and victimization, we also examined the patterns of relations with a wide variety of other psychosocial constructs relevant to bullying research. Thus, for example, students rated their attitudes toward bullies and victims and indicated the role that they would assume if confronted with a situation involving bullying. Whereas some of these relations were in the expected direction, there were also some disconcerting surprises. As expected, bullying was positively related to pro-bully attitudes and negatively related to pro-victim attitudes. We were surprised to find, however, that victimization was also positively related to higher pro-bully attitudes but not to higher pro-victim attitudes. These results suggest that victims have negative feelings toward themselves (as victims) or, perhaps, that they identify themselves more as being bullies than victims or even aspire to become bullies themselves. It is important to note that both bullies and—to a lesser extent—victims indicated that if confronted by a bullying situation, they would be more likely to actively or passively reinforce the bullying behaviors than to ignore the situation or to be an advocate for the victim. These results are also consistent with our reciprocal effects model in which bullies become victims and victims become bullies over time. Victims and bullies are clearly much more alike than they are different. Overall, we were struck by the similarity between the patterns of responses of bullying and victimization with a wide range of psychological constructs. Thus, for example, bullying and victimization were related to endorsing pro-bullying behaviors, actively or passively reinforcing bullying, using avoidance coping strategies, having an external locus of control, experiencing difficulties in controlling anger, being depressed, and having negative self-concepts in most areas. Also, bullying and victimization were negatively related to ignoring a bullying situation, using positive coping strategies (problem solving or seeking social support), and having an internal locus of control.

There were, however, some striking differences between patterns of relations with bullying and victimization. Although both bullying and victimization were negatively related to effective anger management, bullying was related to externalizing anger while victimization was related to internalizing anger. Although both bullying and victimization were positively related to depression and negatively related to self-esteem, clear differences were found in patterns of relations with several of the specific components of self-concept. For victimization, there were reasonably consistent negative relations with different areas of self-concept, although their self-concepts are particularly low for same-sex relationships and self-concepts of emotional stability. For bullying, the patterns were more differentiated, with little or no relations with physical, peer relationships, and emotional-stability self-concepts. For bullying, the most negative relation was for the moral factor of honesty and trustworthiness. Particularly interesting is the juxtaposition between same-sex and opposite-sex self-concepts in relation to bullying and victimization. Opposite-sex relationships self-concept is the one component of self-concept that was positively related to bullying and is clearly different from the relations with same-sex self-concept (which had a slightly negative correlation with bullying). This suggests that bullying might be a strategy that some students used to enhance their self-perceived status with the opposite sex. Victimization is most negatively related to same-sex relationship self-concepts, much more so than for opposite-sex factor (that is somewhat negative). Because the bullying usually is perpetrated by someone of the same sex, it is this facet of self-concept where the victim is particularly likely to suffer. Although it is typically the physical subdomain of bullying and victimization that receives the most attention, our results suggest the social and verbal subdomains are more negatively associated with many psychosocial indicators of psychological health.

The complex patterns of relations between the bullying and victimization factors and the diverse set of correlates (background variables and psychosocial constructs) provide strong support for the convergent and discriminant validity in relation to both domains (bullying and victimization) and subdomains (verbal, social, and physical), as well as the usefulness of ESEM in this specific area of research as well as for educational psychology research more generally.

MTMM support for convergent and discriminant validity. The MTMM design is widely used to evaluate support for convergent and discriminant validity and is routinely applied to evaluation of responses from new and existing instruments. Using time (the three waves of data) as the method factor, we showed good support for the discriminant and convergent validity of the six bully and victim factors. However, several features of these analyses require further consideration. Although, not surprisingly, there was very good support for the differentiation of three bully factors from three victim factors, evaluation of support for discriminant validity within the three bully factors and within the three victim factors was more problematic. In particular, using the traditional ICM–CFA approach, we found little or no support for this aspect of discriminant validity. In marked contrast, there was good support for this aspect of discriminant validity on the basis of ESEM approach. These MTMM analyses—perhaps more than any other aspect of the present investigation—demonstrated the importance of the ESEM approach used here, particularly in relation to tests of discriminant validity of subdomains of bullying and victimization.

Directional ordering of bullying and victimization. A growing body of literature shows that bullying and victimization are not antithetical—that the two tend to be positively correlated. Some persons may participate in bullying behaviors as well as experience victimization by others who take on such behaviors. Here we pursued the more complicated question of how bullying

and being bullied on one occasion influences subsequent bullying and victimization. Correlations based on a single wave of data are heuristic but provide no clear basis for concluding either that bullying leads to victimization or that being victimized leads to bullying. In order to disentangle the directionality of the relationships between these constructs, we considered fully latent autoregressive cross-lagged panel models of bullying and victimization responses collected on three occasions as an extension of the ESEM approach. With this approach, we tested the hypothesis that bullying and victimization are reciprocally related such that they mutually reinforce each other over time. Although there is a long history of the use of such models in other areas of research (e.g., Marsh, Byrne, & Yeung, 1999), we are not aware of prior applications in bullying research (but see Ostrov, in press, as well as discussion of the need for such approaches by Little et al., 2007).

Particularly for the physical subdomains of both bullying and victimization, there was evidence for a reciprocal pattern of relations—prior physical bullying leads to subsequent physical victimization, and prior physical victimization leads to subsequent physical bullying. These results illustrate that the bully and victim roles might be mutually reinforcing and also contribute to the growing recognition of the inappropriateness of treating bullies and victims as mutually exclusive groups or as opposite endpoints of a bipolar construct. These results have important implications for better understanding the nature of bullies and victims, for interventions to decrease bullying in schools, and for future research.

While the methodological approach used here (a combination of ESEM and multiwave/multivariable models of longitudinal data) is ideally suited for evaluation of this issue, the time frame (1 academic year) may not have been sufficiently long for us to fully evaluate the long-term development of these patterns of relations between bullying and victimization. This aspect of our research needs to be tested over a longer time frame in the context of diverse participant groups, settings, and contexts to further test the proposed reciprocal relations between bullying and victimization.

Methodological implications, limitations, and directions for future research. The ESEM approach used here is a particularly useful contribution to bullying and victimization research and to educational psychology research more generally. Although ESEM has only recently been made available to applied researchers, there is growing evidence that it provides a useful integration of traditional EFA, CFA, and SEM approaches. Many widely used educational psychology instruments have apparently well-defined factor structures based on EFA but do not provide an acceptable fit when evaluated with traditional ICM–CFA approaches. Faced with this situation, many applied researchers forgo the advantages of ICM–CFA. In such cases, ESEM might be ideally suited in that many of the advantages of ICM–CFA are available in ESEM. Furthermore, as demonstrated here, there are potential advantages to the use of ESEM in some circumstances even when a more parsimonious ICM–CFA model provides a reasonable fit to the data. In particular, correlations among ICM–CFA factors are likely to be positively biased–perhaps substantially–unless all the nontarget factor loadings are very close to zero.

Here we demonstrated the flexibility of ESEM in a variety of applications that could not readily be accomplished with EFA. Starting with the initial evaluation of the APRI factor structure, ESEM produced parameter estimates (with standard errors), goodness-of-fit statistics, modification indexes, and the potential to impose more complicated error structures (e.g., correlated uniquenesses). Applying the detailed taxonomy of invariance tests, we were able to test configural, weak, strong, and strict models of invariance of the factor structure over gender, year in school groups, and—for longitudinal data—over time. These tests are a critical aspect of the evaluation of potential measurement instruments and essential in terms of validating interpretations based on mean-level comparisons between groups or over time. In particular, problems associated with differential item functioning (a lack of strong measurement invariance) are a major focus in the evaluation of standardized achievement tests and selection instruments but have been largely ignored in bullying research (as well as many other areas of applied psychological research; Card & Little, 2007; Marsh et al., 2009, 2010). Also, with ESEM, it is possible to evaluate potentially complex models of measurement error such as demonstrated with the correlated uniquenesses that are recommended components of longitudinal studies. The application of ESEM to evaluate convergent and discriminant validity on the basis of MTMM design and causal ordering based on multivariate panel designs are apparently new applications of ESEM but also demonstrate its flexibility.

What are the consequences—the theoretical and practice implications—of not applying ESEM and evaluating cross-loadings, instead of simply relying on traditional ICM–CFAs? In the development of instruments, it is important to evaluate cross-loadings as one basis for selecting the best items and more fully understand the meaning of the factors being measured. However, even for well-established instruments that are widely used (e.g., the NEO instrument used to measure big-five personality factors, Marsh et al., 2010; or the Students' Evaluation of Education Quality instrument used to measure university students' evaluations of teaching effectiveness, Marsh et al., 2009), this pursuit is important. It will often be the case—even with widely used instruments like those used in applied research—that the fit of the ICM–CFA model does not meet acceptable standards of fit to the model (see Marsh, Hau & Grayson, 2005; Marsh, Hau & Wen, 2004;), while the ESEM model does. However, an unacceptable goodness of fit is only an indicator that something is wrong in terms of the model. In the case of comparing ESEM and ICM–CFA models, it is often clear that the problem is with the ICM–CFA constraint of nontarget loadings to be zero. Although this may or may not have important implications for identifying the factors and identifying poor items, it clearly has important implications for the size of factor correlations and—ultimately—support for the discriminant validity of the factors and their usefulness in situations where discrimination among the factors is important. The present investigation provides an important caveat to this concern in that the fit of our ICM–CFA model was apparently acceptable, although the fit of the ESEM model was better. Even in this situation in which the fit of the ICM–CFA model was apparently acceptable, ignoring the relatively modest nontarget loadings meant that there was little or no meaningful distinction among the three (verbal, social, physical) facets of bullying and among the corresponding three facets of victimization factors. Furthermore, from our review of this literature, it is apparent that no other instruments in this substantial literature of applied research have been successful in establishing this distinction. These results are clearly important for bully/victimization research but have obvious implications for all ap-

plied research based on psychological measures. We also note that the use of simple scale scores is even more problematic in that, as CFA, it implicitly assumes that there are no cross-loadings (i.e., each item contributes to only one scale) but also because it assumes that all items are equally weighted and fails to take into account measurement error (two points that are usually taken into account in CFAs as well as in ESEMs). Hence, we recommend that ESEM should be considered—at least in preliminary analyses—whenever applied researchers conduct ICM-CFAs.

All ESEM analyses conducted in the present investigation could also have been realized within traditional ICM–CFAs models. In this respect, we present the ESEM model as a viable alternative to the ICM–CFA model but do not argue that the ESEM approach should replace the corresponding ICM–CFA approach in all cases. Hence, ESEM is not a panacea and may not be appropriate in some applications. ESEM should generally be preferred to the ICM–CFA model when the factors are appropriately identified by ESEM, the goodness of fit for ESEM is meaningful better than for ICM–CFA, nontarget loadings are significant, and factor correlations are meaningfully smaller for the ESEM model than the ICM–CFA model. As ICM–CFA is so much more parsimonious, it remains preferable unless there is clear evidence of the superiority of ESEM (better fit, more appropriate parameter estimates). It is, of course, relevant to point out that the ESEM factors might not be consistent with the a priori design of the instrument while the CFA factors are forced to be consistent with it. Although clearly not the case here, this might call into question the appropriateness of the a priori design of the instrument—particularly if the fit of the ICM–CFA model was not particularly good.

Historically, factor analysis was routinely pursued with traditional EFA approaches but was largely superseded by CFA approaches that are much more powerful. However, there are practical limitations with the application of CFA identified here that render dubious its routine application in much applied research. However, ESEM—the integration of EFA, CFA, and SEM—provides researchers with considerable flexibility to address substantively important issues such as those raised here when the traditional ICM–CFA models might not be appropriate. Because ESEM is a new statistical tool, "best practice" will have to evolve with experience. Nevertheless, results of the present investigation provide considerable promise for the application of ESEM for bullying and victimization studies and for educational psychology research more generally (also see Asparouhov & Muthén, 2009; Marsh et al., 2009, 2010).

References

Amirkhan, J. H. (1990). A factor analytically derived measure of coping: The coping strategy indicator. *Journal of Personality and Social Psychology, 59,* 1066–1074. doi:10.1037/0022-3514.59.5.1066

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40,* 57–87. doi:10.1111/1469-7610.00424

Archer, J. (2004). Sex differences in aggression in real world settings: A meta-analytic review. *Review of General Psychology, 8,* 291–322. doi:10.1037/1089-2680.8.4.291

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

Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior, 18,* 117–127. doi:10.1002/1098-2337(1992)18:2<117::AID-AB2480180205>3.0.CO;2-3

Bosworth, K., Espelage, D. L., & Simon, T. R. (1999). Factors associated with bullying behavior in middle school students. *Journal of Early Adolescence, 19,* 341–362. doi:10.1177/0272431699019003003

Boulton, M. J., & Smith, P. K. (1994). Bully/victim problems in middle-school children: Stability, self-perceived competence, peer perceptions and peer acceptance. *British Journal of Developmental Psychology, 12,* 315–329.

Boulton, M. J., & Underwood, K. (1992). Bully/victim problems among middle school children. *British Journal of Educational Psychology, 62,* 73–87.

Boyle, G. J. (1994). Self-Description Questionnaire II: A review. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques: Vol. 10* (pp. 632–643). Kansas City, MO: Test Corp. of America.

Byrne, B. (1996). *Measuring self-concept across the life span: Issues and instrumentation.* Washington, DC: American Psychological Association. doi:10.1037/10197-000

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. doi:10.1037/h0046016

Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research, 2,* 409–426. doi:10.1207/s15327906mbr0204_1

Card, N. A., & Hodges, E. V. E. (2008). Peer victimization among school children: Correlations, causes, consequences, and considerations in assessment and intervention. *School Psychology Quarterly, 23,* 451–461. doi:10.1037/a0012769

Card, N. A., & Little, T. D., (2007). Studying aggression with structural equation modeling. In D. Flannery, A. Vazsonyi, & I. Waldman (Eds.), *The Cambridge Handbook of Violent Behavior* (pp. 727–739). New York: Cambridge University Press.

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 intercorrelations, gender differences, and relations to maladjustment. *Child Development, 79,* 1185–1229. doi:10.1111/j.1467-8624.2008.01184.x

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

Cheung, G. W., & Rensvold, R. B. (2001). The effects of model parsimony and sampling error on the fit of structural equation models. *Organizational Research Methods, 4,* 236–264. doi:10.1177/109442810143004

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31,* 33–42. doi:10.1007/BF02289455

Cohen, J. (1968). Multiple regression as a general data analytic system. *Psychological Bulletin, 70*(6), 426–443.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Crick, N. R., Bigbee, M. A., & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development, 67,* 1003–1014. doi:10.2307/1131876

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., Nelson, D. A., Morales, J. R., Cullerton-Sen, C., Casas, J. F., & Hickman, S. E. (2001). Relational victimization in childhood and adolescence: I hurt you through the grapevine. In J. Juvonen, A. Nishina, & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 196–214). New York: Guilford.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory

factor model. *Structural Equation Modeling, 16,* 295–314. doi:10.1080/10705510902751416

Due, P., Holstein, B. E., Lynch, J., Diderichsen, F., Gabnain, S. N., Scheidt, P., Currie, C., & The Health Behaviour in School-Aged Children Bullying Working Group. (2005). Bullying and symptoms among school-aged children: International comparative cross sectional study in 28 countries. *European Journal of Public Health, 15,* 128–132. doi:10.1093/eurpub/cki105

Enders, C. (2001). A primer on maximum likelihood algorithms for use with missing data. *Structural Equation Modeling, 8,* 128–141. doi:10.1207/S15328007SEM0801_7

Feshbach, S., & Zagrodzka, J. (Eds.). (1997). *Aggression: Biological, developmental, and social perspectives.* New York: Plenum Press.

Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2007). Polyvictimization and trauma in a national longitudinal cohort. *Development and Psychopathology, 19,* 149–166. doi:10.1017/S0954579407070083

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576. doi:10.1146/annurev.psych.58.110405.085530

Gumpel, T. P. (2008). Behavioral disorders in the school participant roles and sub-roles in three types of school violence. *Journal of Emotional and Behavioral Disorders, 16,* 145–162. doi:10.1177/10634266073108 46

Harachi, T. W., Catalano, R. F., & Hawkins, D. J. (1999). United States. In P. K. Smith, Y. Morita, J. Junger-Tas, D. Olweus, R. F. Catalano, & P. T. Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp.279–295). London, England: Routledge.

Hattie, J. (1992). *Self-concept.* Hillsdale, NJ: Lawrence Erlbaum.

Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years' research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry, 41,* 441–455. doi:10.1111/1469-7610.00629

Healey, J. (2001). Bullying and youth violence. *Issues in Society, Vol. 154*. Rozelle, New South Wales, Australia: Spinney.

Hussein, M. H. (2010). The Peer Interaction in Primary School Questionnaire: Testing for measurement equivalence and latent mean differences in bullying between gender in Egypt, Saudi Arabia, and the USA. *Social Psychology of Education, 13,* 57–76. doi:10.1007/s11218-009-9098-y

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal investigations. In J. R. Nesselroade & B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–351). New York, NY: Academic Press.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt.

Kovacs, M. (1992). *Children's Depression Inventory (CDI) manual*: New York, NY: Multi Health Systems.

Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *Journal of Child Psychology and Psychiatry, 39,* 47–63. doi:10.1017/S0021963097001765

Lagerspetz, K. M. J., Björkqvist, K., Berts, M., & King, E. (1982). Group aggression among school children in three schools. *Scandinavian Journal of Psychology, 23,* 45–52. doi:10.1111/j.1467-9450.1982.tb00412.x

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

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31,* 357–365. doi:10.1177/0165025407077757

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

Marsh, H. W. (1988). Multitrait–multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement, and evaluation: An international handbook* (pp. 570–580). Oxford, England: Pergamon Press.

Marsh, H. W. (1990). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph.* San Antonio, TX: Psychological Corp.

Marsh, H. W. (1993a). Multitrait-multimethod analyses: Inferring each trait/method combination with multiple indicators. *Applied Measurement in Education, 6,* 49–81. doi:10.1207/s15324818ame0601_4

Marsh, H. W. (1993b). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement, 30,* 157–183. doi:10.1111/j.1745-3984.1993.tb01072.x

Marsh, H. W. (2007a). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774–798). New York, NY: Wiley.

Marsh, H. W. (2007b). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology.* Leicester, UK: British Psychological Society.

Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315–353). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410. doi:10.1037/0033-2909.103.3.391

Marsh, H. W., Byrne, B. M., & Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist, 34,* 155–167. doi:10.1207/s15326985ep3403_2

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1,* 133–163. doi:10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Ellis, L., Parada, L., Richards, G., & Heubeck, B. G. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment, 17,* 81–102. doi:10.1037/1040-3590.17.1.81

Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. *Structural Equation Modeling, 1,* 317–359. doi:10.1080/10705519409539984

Marsh, H. W., & Grayson, D. (1995). Latent-variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

Marsh, H. W., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). Oxford, UK: Wiley.

Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education, 64,* 364–390.

Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology, 32,* 151–170. doi:10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness-of-fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Psychometrics. A festschrift to Roderick P. McDonald* (pp. 275–340). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules:

Comment on hypothesis testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341. doi:10.1207/s15328007sem1103_2

Marsh, H. W., & Hocevar, D. (1988). A new procedure for analysis of multitrait–multimethod data: An application of second order confirmatory factor analysis. *Journal of Applied Psychology, 73,* 107–117. doi:10.1037/0021-9010.73.1.107

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471–491. doi:10.1037/a0019227

Marsh, H. W., Lüdtke, O., Robitzsch, A., Nagengast, B., Morin, A. J. S., & Trautwein, U. (2011). Two wrongs do not make a right: Camouflaging misfit with item-parcels in CFA models. Manuscript submitted for publication.

Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology, 32,* 438–482.

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

Marsh, H. W., & O'Mara, A. (2010). Long-term total negative effects of school-average ability on diverse educational outcomes: Direct and indirect effects of the big-fish–little-pond effect. *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology, 24,* 51–72.

Marsh, H. W., Parada, R. H., Craven, R. G., & Finger, L. (2004). In the looking glass: A reciprocal effect model elucidating the complex nature of bullying, psychological determinants, and the central role of self-concept. In C. E. Sanders & G. D. Phye (Eds.), *Bullying: Implications for the classroom* (pp. 63–109). San Diego, CA: Elsevier.

Marsh, H. W., Parada, R. H., Yeung, A. S., & Healey, J. (2001). Aggressive school troublemakers and victims: A longitudinal model examining the pivotal role of self-concept. *Journal of Educational Psychology, 93,* 411–419. doi:10.1037/0022-0663.93.2.411

Mayer, M. J., & Cornell, D. G. (2010). New perspectives on school safety and violence prevention. *Educational Researcher, 39,* 5–6. doi:10.3102/0013189X09356778

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika, 29,* 187–206. doi:10.1007/BF02289700

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543. doi:10.1007/BF02294825

Meredith, W., & Teresi, J. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(Suppl. 3), S69–S77. doi:10.1097/01.mlr.0000245438.73837.89

Muthén, L. K., & Muthén, B. (2008). *Mplus user's guide.* Los Angeles CA: Muthén & Muthén.

Mynard, H., & Joseph, S. (2000). Development of the multidimensional peer-victimization scale. *Aggressive Behavior, 26,* 169–178. doi:10.1002/(SICI)1098-2337(2000)26:2<169::AID-AB3>3.0.CO;2-A

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, J. W., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association, 285,* 2094–2100. doi:10.1001/jama.285.16.2094

Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 411–448). Hillsdale, NJ: Erlbaum.

Olweus, D. (1993). Bully/victim problems among schoolchildren: Long-term consequences and an effective intervention program. In S. Hodgins (Ed.), *Mental disorder and crime* (pp. 317–349). Newbury Park, CA: Sage.

Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education, 12,* 495–510. doi:10.1007/BF03172807

Ostrov, J. M. (in press). Prospective associations between peer victimization and aggression. *Child Development.*

Ostrov, J. M., & Godleski, S. A. (2010). Toward an integrated gender-linked model of aggression subtypes in early and middle childhood. *Psychological Review, 117,* 233–242. doi:10.1037/a0018070

Parada, R. H. (2000). *Adolescent Peer Relations Instrument: A theoretical and empirical basis for the measurement of participant roles in bullying and victimization of adolescence. An interim test manual and a research monograph: A test manual*. Penrith, New South Wales, Australia: University of Western Sydney, Self-Concept Enhancement and Learning Facilitation (SELF) Research Centre, Publication Unit.

Parada, R. H. (2002). Beyond bullying secondary schools program: Consultant's handbook. Penrith, New South Wales, Australia: University of Western Sydney, Self-Concept Enhancement and Learning Facilitation (SELF) Research Centre, Publication Unit.

Parada, R. H. (2006). *School bullying: Psychosocial determinants and effective intervention.* Unpublished doctoral dissertation, University of Western Sydney, Penrith, New South Wales, Australia.

Parada, R. H., Craven, R. G., & Marsh, H. W. (2008). The beyond bullying secondary program: An innovative program empowering teachers to counteract bullying in schools. *Self-processes, learning, and enabling human potential: Dynamic new approaches* (Vol.3, pp. 373–426). Charlotte, NC: Information Age.

Parada, R. H., Marsh, H. W., & Craven, G. R. (2005, November). *There and back again from bully to victim and victim to bully: A reciprocal effects model of bullying behaviours in schools.* Paper presented at the Australian Association for Research in Education International, Parramatta, New South Wales, Australia.

Rigby, K. (1996). *Bullying in schools—and what to do about it.* Melbourne, Australia: ACER.

Rigby, K., & Slee, P. T. (1999). Suicidal ideation among adolescent school children, involvement in bully–victim problems, and perceived social support. *Suicide and Life-Threatening Behavior, 29,* 119–130.

Rivers, I., & Smith, P. K. (1994). Types of bullying behavior and their correlates. *Aggressive Behavior, 20,* 359–368. doi:10.1002/1098-2337(1994)20:5<359::AID-AB2480200503>3.0.CO;2-J

Roland, E., & Idsøe, T. (2001). Aggression and bullying. *Aggressive Behavior, 27,* 446–462. doi:10.1002/ab.1029

Salmivalli, C., Kaukiainen, A., & Lagerspetz, K. M. J. (2000). Aggression and sociometric status among peers: Do gender and type of aggression matter? *Scandinavian Journal of Psychology, 41,* 17–24. doi:10.1111/1467-9450.00166

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45,* 73–103. doi:10.1080/00273170903504810

Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. R. van de Vijver, D. A., van Hemert, & Y. H. Poortinga (Eds.), *Multilevel analysis of individuals and cultures.* (pp. 93–119). New York, NY: Taylor & Francis.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research, 46,* 407–441.

Smith, P. K., Morita, Y., Junger-Tas, J., Olweus, D., Catalano, R. F., & Slee, P. (Eds.). (1999). *The nature of school bullying: A cross-national perspective.* London, England: Routledge.

Smith, P. K., & Sharp, S. (Eds.). (1994). *School bullying: Insights and perspectives.* London, England: Routledge. doi:10.4324/9780203425497

Sullivan, K. (2000). *The anti-bullying handbook.* Auckland: New Zealand: Oxford University Press.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and "theory of mind": A critique of the "social skills deficit" view of anti-social behavior. *Social Development, 8,* 117–127. doi:10.1111/1467-9507.00083

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying? Linking research to educational practice. *Educational Researcher, 39,* 38–47. doi:10.3102/0013189X09357622

Underwood, M. K. (2003). *Social aggression among girls.* New York, NY: Guilford.

Vaillancourt, T., Brendgen, M., Boivin, M., & Tremblay, R. E. (2003). A longitudinal confirmatory factor analysis of indirect and physical aggression: Evidence of two factors over time? *Child Development, 74,* 1628–1638. doi:10.1046/j.1467-8624.2003.00628.x

Wolke, D., Woods, S., Bloomfield, L., & Karstadt, L. (2000). The association between direct and relational bullying and behavior problems among primary school children. *Journal of Child Psychology and Psychiatry, 41,* 989–1002. doi:10.1111/1469-7610.00687

Wylie, R. C. (1989). *Measures of self-concept.* Lincoln: University of Nebraska.

Appendix A

Adolescent Peer Relations Instrument: Bully/Target (*A*PRI-*BT*)

© 2010 **Roberto H Parada, Herbert W Marsh & Rhonda Craven.**

SECTION A

Since you have been at this school THIS YEAR how often HAVE YOU done any of the following things to a STUDENT (or students) at this school. CIRCLE THE NUMBER THAT IS CLOSEST TO YOUR ANSWER.

	Never	Some-times	Once or twice a month	Once a week	Several times a week	Every day
In the past year at this school I . . .						
1. Teased them by saying things to them	1	2	3	4	5	6
2. Pushed or shoved a student	1	2	3	4	5	6
3. Made rude remarks at a student	1	2	3	4	5	6
4. Got my friends to turn against a student	1	2	3	4	5	6
5. Made jokes about a student	1	2	3	4	5	6
6. Crashed into a student on purpose as they walked by	1	2	3	4	5	6
7. Picked on a student by swearing at them	1	2	3	4	5	6
8. Told my friends things about a student to get them into trouble	1	2	3	4	5	6
9. Got into a physical fight with a student because I didn't like them	1	2	3	4	5	6
10. Said things about their looks they didn't like	1	2	3	4	5	6
11. Got other students to start a rumour about a student	1	2	3	4	5	6
12. I slapped or punched a student	1	2	3	4	5	6
13. Got other students to ignore a student	1	2	3	4	5	6
14. Made fun of a student by calling them names	1	2	3	4	5	6
15. Threw something at a student to hit them	1	2	3	4	5	6
16. Threatened to physically hurt or harm a student	1	2	3	4	5	6
17. Left them out of activities or games on purpose	1	2	3	4	5	6
18. Kept a student away from me by giving them mean looks	1	2	3	4	5	6

(*Appendices continue*)

SECTION B

Please indicate how often a student (or students) **at this school** has done the following things TO YOU since you have **been at this school this year**. CIRCLE THE NUMBER THAT IS CLOSEST TO YOUR ANSWER

	Never	Some-times	Once or twice a month	Once a week	Several times a week	Every day
In the past year at this school . . .						
1. I was teased by students saying things to me	1	2	3	4	5	6
2. I was pushed or shoved	1	2	3	4	5	6
3. A student wouldn't be friends with me because other people didn't like me	1	2	3	4	5	6
4. A student made rude remarks at me	1	2	3	4	5	6
5. I was hit or kicked hard	1	2	3	4	5	6
6. A student ignored me when they were with their friends	1	2	3	4	5	6
7. Jokes were made up about me	1	2	3	4	5	6
8. Students crashed into me on purpose as they walked by	1	2	3	4	5	6
9. A student got their friends to turn against me	1	2	3	4	5	6
10. My property was damaged on purpose	1	2	3	4	5	6
11. Things were said about my looks I didn't like	1	2	3	4	5	6
12. I wasn't invited to a student's place because other people didn't like me	1	2	3	4	5	6
13. I was ridiculed by students saying things to me	1	2	3	4	5	6
14. A student got students to start a rumour about me	1	2	3	4	5	6
15. Something was thrown at me to hit me	1	2	3	4	5	6
16. I was threatened to be physically hurt or harmed	1	2	3	4	5	6
17. I was left out of activities, games on purpose	1	2	3	4	5	6
18. I was called names I didn't like	1	2	3	4	5	6

© 2010 Roberto H Parada; Herbert W Marsh & Rhonda Craven. The authors grant permission for the free use of this instrument for research and non-commercial purposes only. Please contact the first author for further information: parada_roberto@hotmail.com.

Adolescent Peer Relations Instrument: Bully/Target (*A*PRI-*BT*)
© 2010 Roberto H Parada, Herbert W Marsh & Rhonda Craven.

SCORING KEY

Section A are the bullying items Section B are the Victim(Target) items. Items are labled as relating to Social Bullying, Verbal Bullying, and Physical Bullying. Scoring is achieved by simply adding the items up. Any student who scores less than 36 (or 18 in each scale) has never been bullied or has never bullied others. There is no cut off scores for this instrument. Each student receives both a total Victim and Bully score as well as six scores for each of the scales. A score = or less than 6 on any scale means never been bullied or has never bullied others in that particular way.

Citing this instrument: Please cite this article and the test manual:

Parada, R. (2000). *Adolescent Peer Relations Instrument: A theoretical and empirical basis for the measurement of participant roles in bullying and victimisation of adolescence: An interim test manual and a research monograph: A test manual.* Publication Unit, Self-concept Enhancement and Learning Facilitation (SELF) Research Centre, University of Western Sydney.

(*Appendices continue*)

SECTION A

	Scale
In the past year at this school I . . .	
1. Teased them by saying things to them	Verbal
2. Pushed or shoved a student	Physical
3. Made rude remarks at a student	Verbal
4. Got my friends to turn against a student	Social
5. Made jokes about a student	Verbal
6. Crashed into a student on purpose as they walked by	Physical
7. Picked on a student by swearing at them	Verbal
8. Told my friends things about a student to get them into trouble	Social
9. Got into a physical fight with a student because I didn't like them	Physical
10. Said things about their looks they didn't like	Verbal
11. Got other students to start a rumour about a student	Social
12. I slapped or punched a student	Physical
13. Got other students to ignore a student	Social
14. Made fun of a student by calling them names	Verbal
15. Threw something at a student to hit them	Physical
16. Threatened to physically hurt or harm a student	Physical
17. Left them out of activities or games on purpose	Social
18. Kept a student away from me by giving them mean looks	Social

SECTION B

	Scale
In the past year at this school I . . .	
1. I was teased by students saying things to me	Verbal
2. I was pushed or shoved	Physical
3. A student wouldn't be friends with me because other people didn't like me	Social
4. A student made rude remarks at me	Verbal
5. I was hit or kicked hard	Physical
6. A student ignored me when they were with their friends	Social
7. Jokes were made up about me	Verbal
8. Students crashed into me on purpose as they walked by	Physical
9. A student got their friends to turn against me	Social
10. My property was damaged on purpose	Physical
11. Things were said about my looks I didn't like	Verbal
12. I wasn't invited to a student's place because other people didn't like me	Social
13. I was ridiculed by students saying things to me	Verbal
14. A student got students to start a rumour about me	Social
15. Something was thrown at me to hit me	Physical
16. I was threatened to be physically hurt or harmed	Physical
17. I was left out of activities, games on purpose	Social
18. I was called names I didn't like	Verbal

© 2010 Roberto H Parada; Herbert W Marsh & Rhonda Craven. The authors grant permission for the free use of this instrument for research and non-commercial purposes only. Please contact the first author for further information: parada_roberto@hotmail.com.

(*Appendices continue*)

Appendix B

Scales Considered in the Present Investigation: Description, Reliability, and Length

Instrument Description

The Adolescent Peer Relations Instruments (APRI) are multidimensional measures, which were specifically developed for this study to measure interpersonal relationships among high school students (Parada, 2000). The APRI–BT measures bullying and victimization in three subdomains (verbal, physical and social). In the first section, students are asked to state how often on a 6-point Likert scale (from 1 = *never* to 6 = *everyday*) do they engage in a series of behaviors against other students. In the second section, the students are asked how often these behaviors have occurred to them over the past year. The APRI–PR measures students participant roles (active reinforcer, passive reinforcer, victim advocate, and disregard/ignore) when they witness a bullying situation. Students are asked to rate how true these reactions are to what they would do on a 6-point Likert scale (from 1= *false* to 6 = *true*). The APRI–A measures students' bullying attitudes (pro-bully, pro-victim) by asking them to rate on a 6-point Likert scale how much they agree or disagree (from 1 = *completely disagree* to 6 = *agree*) with a series of statements reflecting pro-victims or pro-bully attitudes.

A coping strategies questionnaire modeled on Amirkhan's (1990) work was used to measure three coping strategies commonly used by adolescents when faced with difficulties: (a) active problem solving, (b) seeking social support, or (c) problem avoidance. A stem statement "When I have a problem . . . " preceded 15 items describing various ways of dealing with problems. Students rated on a 6-point Likert scale (1 = *never* to 6 = *always*) how they react when they face difficulties in general.

A locus of control (LOC) scale was specifically developed to measure the extent to which students attribute their success or failures in day-to-day living to internal (e.g., effort) or external (e.g., luck) factors. Students were asked to show how much they agreed or disagreed with eight statements about life on a 6-point Likert scale (1 = *completely disagree* to 6 = *agree*).

The Anger Expression Index–Adolescent consists of 12 items developed to measure three components of anger expression that are relevant to this study: (a) internalizing feelings of anger, (b) externalizing feelings of anger (by lashing out for example), and (c) controlling feelings of anger in appropriate ways. A stem statement "When I am angry . . . " precedes items describing various ways of reacting to anger. Students were asked to rate on a 6-point Likert scale (1 = *never* to 6 = *always*) how they react when they are angry.

The short version of the Child Depression Inventory (Kovacs, 1992), one of the most widely used assessment tools for depression in both clinical and nonclinical populations, was used to assess depressive symptoms in the students. Each of the 10 items consists of three statements expressing different levels of a depressive symptom. Students were asked to indicate which statement is the truest for them by placing a mark next to their choice.

The self-concept measure was the Self-Description Questionnaire II–Short Form (Marsh, Ellis, Parada, Richards, & Heubeck, 2005; also see Marsh, 1990, 2007b), an Australian instrument that is widely regarded internationally to be the best available multidimensional self-concept instrument for this age group (e.g., Boyle, 1994; Byrne, 1996; Hattie, 1992). Participants respond to each of the 51 items on a 6 point Likert scale (1= *false* to 6 = *true*), resulting in 11 distinct self-concept factors (three academic: math, verbal, academic; 8 nonacademic: physical, appearance, opposite-sex relationships, same-sex relationships, honesty and trustworthiness, emotional stability, and global self-esteem).

In the following section, the coefficient alpha estimates of reliability for each scale are presented for responses at Times 1, 2, and 3, respectively (in parentheses), along with the number of items and sample items from the scales.

Instrument Reliabilities and Sample Items

Adolescent Peer Relations Instruments (APRI) Bullying Scales

Bully–verbal (.89, .90, .92). Six items. Sample item: "Teased them by saying things to them."

Bully–social (.82, .86, .90). Six items. Sample item: "Got my friends to turn against a student."

Bully–physical (.85, .87, .90). Six items. Sample item: "Got into a physical fight with a student because I didn't like them."

APRI Victimization (Target of Bullying) Scales

Victim–verbal (.92, .92, .93). Six items. Sample item: "I was teased by students saying things to me."

Victim–social (.87, .91, .92). Six items. Sample item: "A student wouldn't be friends with me because other people didn't like me."

Victim–physical (.89, .89, .92). Six items. Sample item: "I was threatened to be physically hurt or harmed."

APRI–PR (Participation Roles)

Active reinforcer (.78, .86, .87). Six items. Sample item: "I would join in myself."

Passive reinforcer (.88, .90, .91). Six items. Sample item: "I would stay to watch what happens."

Ignore (.87, .89, .91). Six items. Sample item: "I would pay no attention to it."

Victim advocate (.89, .91, .92). Six items. Sample item: "I would get my friends to help me stop it."

APRI–A (Bully Attitudes)

Pro-bully (.64, .68, .74). Six items. Sample item: "It's OK to bully others if others are doing it."

(*Appendices continue*)

Pro-victim (.68, .74, .79). Six items. Sample item: "People who are bullied deserve our help."

Coping Style

Avoidance (.75, .79, .83). Six items. Sample item: "I avoid the problem by spending more time alone."

Problem solving (.85, .87, .89). Five items. Sample item: "I develop a plan about how to solve the problem before doing anything."

Seek support (.90, .91, .92). Four items. Sample item: "I go to a friend for advice on how to solve the problem."

Locus of Control (LOC) Scale

Internal (.75, .81, .85). Four items. Sample item: "My own efforts and actions are what will determine my future."

External (.71, .75, .78). Four items. Sample item: "External things mostly control my life."

Anger Expression Index–Adolescent

Control (.85, .87, .88). Four items. Sample item: "I stay steady and in control."

Internalize (.67, .69, .75). Four items. Sample item: "No one can tell but I am furious inside."

Externalize (.66, .68, .71). Four items. Sample item: "I let people see just how angry I am."

Childhood Depression Inventory–Short Form

Depression (.83, .85, .76). Ten items. Sample item (respondents are asked to mark the sentence that describes their feelings during the previous 2 weeks): "I am sad once in a while/I am sad many times/I am sad all the time."

Self-Description Questionnaire II–Short Form: Self-Concept

Physical (.83, .84, .84). Four items. Sample item: "I enjoy things like sports, gym, and dance."

Appearance (.87, .88, .89). Four items. Sample item: "I have a nice-looking face."

Opposite-sex relations (.85, .84, .85). Four items. Sample item: "I am not very popular with members of the opposite sex."

Same-sex relations (.79, .82, .83). Five items. Sample item: "It is difficult to make friends with members of my own sex."

Parent relations (.84, .86, .85). Five items. Sample item: "I get along well with my parents."

Honesty and trustworthiness (.79, .81, .81). Six items. Sample item: "I am honest."

Emotional (.81, .83, .85). Five items. Sample item: "I worry more than I need to."

General (.82, .84, .85). Six items. Sample item: "Overall, I have a lot to be proud of."

Math (.90, .90, .90). Four items. Sample item: "Mathematics is one of my best subjects."

Verbal (.89, .90, .91). Five items. Sample item: "I am hopeless in English classes."

School (.84, .86, .86). Four items. Sample item: "I get bad marks in most school subjects."

Appendix C

Scale Items for Instruments in the Present Investigation

Included here are the items for the Self-Description Questionnaire II Short instrument and for new scales that were developed specifically for use in the present investigation (also see Appendix B; for further detail about the development of new scales, see Parada, 2006).

Table C1
Self-Description Questionnaire II Short Form (Marsh, 1990)

Scale description and items
Physical abilities
I enjoy things like sports, gym, and dance
I am good at things like sports, gym, and dance
I am awkward at things like sports, gym, and dance*
I am better than most of my friends at things like sports, gym, and dance
Physical appearance
I have a nice looking face
I am good-looking
Other people think I am good-looking
I have a good-looking body

(*Appendices continue*)

Table C1 (*continued*)

Scale description and items
Same-sex relationships
It is difficult to make friends with members of my own sex*
(If male) I make friends easily with boys; (if female) I make friends easily with girls
Not many people of my own sex like me*
(If male) I do not get along very well with boys; (if female) I do not get along very well with girls
I make friends easily with members of my own sex
Opposite-sex relationships
I am not very popular with members of the opposite sex*
(If female) I make friends easily with boys; (if male) I make friends easily with girls
I have lots of friends of the opposite sex
(If female) I do not get along very well with boys; (if male) I do not get along very well with girls
Honesty/trustworthiness
I am honest
I often tell lies*
I sometimes cheat*
I always tell the truth
I sometimes take things that belong to other people*
I sometimes tell lies to stay out of trouble*
Parent relations
I get along well with my parents
My parents treat me fairly
My parents understand me
I do not like my parents very much*
Emotional stability
I worry more than I need to*
I am a nervous person*
I often feel confused and mixed up*
I get upset easily*
I worry about a lot of things*
Verbal/English
I am hopeless in English classes*
Work in English class is easy for me
English is one of my best subjects
I get good marks in English
I learn quickly in English classes
Mathematics
Mathematics is one of my best subjects
I get good marks in mathematics
I have always done well in mathematics
I do badly in tests in mathematics*
School
I get bad marks in most school subjects*
I learn things quickly in most school subjects
I do well in tests in most school subjects
I am good at most school subjects
General self-esteem
Overall, I have a lot to be proud of
Most things I do, I do well
Overall, most things I do turn out well
I do things as well as most people
If I really try I can do almost anything I want to do
Overall, I am a failure

Note. Items are reverse coded and have been adjusted accordingly for analysis. Items 21 (I make friends easily with boys), 22 (I make friends easily with girls), 43 (I do not get along very well with boys), and 44 (I do not get along very well with girls) were coded according to sex. Scoring adjustments here were made to suit the self-concept domains of opposite-sex relationship and same-sex relationship.

(*Appendices continue*)

Table C2
Adolescent Peer Relations Instrument: Participant Roles (Parada, 2006)

Scale description and items
Active Reinforcer
I would join in myself
I would urge others to join in
I would join in by calling the student being bullied names
I would give suggestions on how to pick on the student
I would join in by laughing at the student being bullied
I would help those who are doing the teasing
Target Advocate
I would get my friends to help me stop it
I would stop it
I would try and protect the student being picked on
I would get help for the student being picked on
I would let someone know who could help stop it
I would take the student being picked on to a safe place
Passive Reinforcer
I would watch and laugh but not join in
I would stay to watch what happens
I would just watch but not join in
I would not join in but call others to come and watch
I would enjoy watching but not join in
I would find it entertaining but not join in
Ignore/Disregard
I would pay no attention to it
I would move away from them
I would ignore it because it's none of my business
I would walk away
I would ignore it
I would mind my own business

Note. Participant role factor were preceded by the stem sentence "If I saw someone being bullied . . ."

Table C3
Adolescent Peer Relations Instrument: Bullying Attitudes (Parada, 2006)

Scale description and items
Pro-bullying
It's OK to bully others if others are doing it
Bullying helps people by making them tougher
Most students who get bullied bring it on themselves
It's OK to bully others to get even
Bullying is OK if done in fun
Other students look up to people who bully others
Pro-target
People who are bullied don't deserve it
Bullying is not OK
Bullying should be stopped
People who are bullied deserve our help
There are no good reasons to bully other students
People who are bullied suffer

(*Appendices continue*)

Table C4
Adolescent Coping Strategy Indicator (Modeled on Amirkan 1990)

Scale description and items
Avoidance
I avoid the problem by spending more time alone
I avoid the problem by watching television more than usual
I avoid the problem by sleeping more than usual
I avoid the problem by pretending that there is no problem
I avoid the problem by staying away from other people
I avoid the problem by wishing that people would leave me alone
Problem solving
I develop a plan about how to solve the problem before doing anything
I set goals for myself to deal with the problem
I make a plan of action about what I will do
I try different ways to solve the problem until I find one that works
I think about what needs to be done
Seek support
I go to a friend for advice on how to solve the problem
I go to a friend to help me feel better
I tell my fears and worries to a friend
I ask my friends to support me

Table C5
Locus of Control Indicator (Parada, 2006)

Scale description and items
External
Other people and events dominate my life
My future is mostly in the hands of other people
Luck and or other people and events control most of my life
External things mostly control my life
Internal
Most good things that happen to me are the result of my own actions
What I do and how I do it will determine my successes in life
If I succeed in life, it will be because of my efforts
My own efforts and actions are what will determine my future

Table C6
Anger Expression Index—Adolescent (Parada, 2006)

Scale description and items
Internalize
No one can tell but I am furious inside
I get upset inside
I boil inside
I'm upset much more than people are aware of
Externalize
I argue with others
I let people see just how angry I am
I don't care if everyone knows
I let it all out
Control
I control my behavior
I control my angry feelings
I keep my cool
I stay steady and in control

Note. Items in the Adolescent Anger Expression Index were preceded by the stem sentence "When I am angry . . . "

Received June 3, 2010
Revision received April 12, 2011
Accepted April 29, 2011 ■