The Judge Specificity of Evaluations of Youth Social Behavior: The Case of Peer Provocation

Melanie A. Dirks, *McGill University*, Teresa A. Treat, *Yale University*, and V. Robin Weersing, *San Diego State University/University of California at San Diego*

Abstract

Increasingly, theorists have suggested that social competence is an evaluative construct that will vary as a function of who is judging behavior. This study examined how two key groups in children's social environments—peers (N = 663, age range = 10.83–15.25 years) and teachers (N = 49)—rated the effectiveness of different behaviors generated by youth in response to physical, verbal, and relational provocation by peers. Results indicated that youth evaluated aggressive responses as more effective than teachers whereas teachers evaluated seeking an explanation as more effective than youth. Both groups judged responses combining assertive and aggressive elements to be more effective than 'pure' aggression. Implications for measurement and intervention models of social competence are discussed.

Keywords: social competence; social skills; behavioral evaluation

Introduction

A significant body of work has been devoted to developing theoretical and measurement models of youth social competence (see Dirks, Treat, & Weersing, 2007a; Rose-Krasnor, 1997). Interest in this topic is driven by at least three factors. Firstly, numerous studies have linked problematic peer relations in childhood to later psychosocial difficulties (see Bierman, 2004; Parker & Asher, 1987). Secondly, social functioning has been implicated in youth psychopathology, including both internalizing (e.g., Rudolph et al., 2000) and externalizing problems (e.g., Dodge, Laird, Lochman, Zelli, & the Conduct Problems Prevention Research Group, 2002). Finally, the 'incompetent' social behaviors themselves may be distressing or harmful to others (e.g., experiencing physical aggression leads to negative emotions for many victims; Kochenderfer-Ladd, 2004). In this article we sought to extend understanding of what constitutes socially competent behavior in a key interpersonal context by examining how peers and teachers evaluated the effectiveness of youth responses to peer provocation.

Correspondence should be addressed to Melanie A. Dirks, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Quebec, Canada H3A 1B1. Email: *melanie.dirks@ mcgill.ca*

Although there is significant heterogeneity in conceptualizations of social compe-tence (Dodge, 1985), there is general agreement that social competence refers to effectiveness in interpersonal interactions (Rose-Krasnor, 1997). Furthermore, researchers have identified four sources of variability in social competence: child, behavior, situation, and judge (Dirks et al., 2007a). In other words, knowing which child enacted which behavior in which situation, as evaluated by which judge, should reveal the most about social competence. To date, most researchers have focused on child-level factors such as physical attractiveness (e.g., Langlois & Stephan, 1977), as well as behavioral correlates of social competence, such as physical aggression (see Ladd, 2005). Although less empirical work has focused on the role of context, the situation specificity of competence is acknowledged increasingly (Rose-Krasnor, 1997).

This study addressed the judge specificity of youth social competence, the fourth dimension underlying this construct. We examined how two key groups in youth's social environments—peers and teachers—evaluated the effectiveness of youth responses to different types of peer provocation, a situation identified as commonly occurring, difficult to manage, and critical (Farrell, Ampy, & Meyer, 1998; Farrell et al., 2006). This work provided detailed information about how peers and teachers, who are both in a position to reward or punish youth behavior, evaluated the range of strategies generated by youth when confronted with physical, verbal, and relational provocation by same-age, same-sex peers.

Inter-rater Discrepancies in Evaluations of Youth Social Competence

Inter-rater discrepancies in evaluations of youth social competence have been well documented. A recent meta-analysis examining the relationships among informants' ratings of children's social competence revealed only modest correlations between teachers and peers, .48; parents and peers, .39; and parents and teachers, .38 (Renk & Phares, 2004). Although there is significant overlap in their ratings, peers, parents, and teachers clearly differed in their perceptions of the target children's competence. In general, researchers have treated such differences as error and collapsed raters' judg-ments to form a single index of competence (see De Los Reyes & Kazdin, 2004). However, variability between judges may reflect at least some valuable signal rather than being attributable entirely to measurement-related noise. For example, some authors have suggested that these discrepancies may reflect raters' experience in different contexts (e.g., school vs. home), across which children's behavior will vary (Achenbach, McConaughy, & Howell, 1987). Judge-specific differences might also reflect useful information about which groups will view a given behavior as effective.

Based on the types of measures used by most researchers, these inter-rater correla-tions of youth social competence can address one of two questions. Many research teams have obtained judgments of children's popularity or acceptance through socio-metric procedures or questionnaires from multiple judges (e.g., Cole, Gondoli, & Peeke, 1998; Connolly & Doyle, 1981). Using these methods, researchers are compar-ing judgments of how well liked a child is. Alternately, researchers have used more specific behavioral nominations (e.g., co-operates or starts fights; Coie & Dodge, 1988) or behavioral rating scales (e.g., child behavior checklist; Schneider & Byrne, 1989). These approaches allow researchers to assess similarities in different judges' perceptions of how often youth engage in specific behaviors. Thus, previous studies have addressed the extent of judges' agreement about *whether* a child is liked or *what*

a child is like (see Parker & Asher, 1987). In general, however, far less is known about differences in ratings of the *competence of actual behaviors*. Addressing this issue requires assessing the effectiveness of different social strategies, as perceived by people actually in children's environments. Such data would build upon the oft-used sociometric and rating scale approaches by helping to determine which behaviors, from whose perspective, contribute to social successes and struggles in important interpersonal contexts.

In the absence of information concerning how children's actions are perceived by the relevant judges in their environments, researchers themselves must decide what constitutes efficacious behavior. For example, when coding youth behavior, investigators have used valenced categories such as 'competent' (e.g., Dodge et al., 2002) or 'effective' (Slaby & Guerra, 1988). However, there is increasing debate about the appropriateness of such judgments. Much of the discussion has centered on aggressive behavior, which Sutton, Smith, and Swettenham (1999) have suggested is undesirable but not necessarily incompetent. Consistent with this viewpoint, some research has demonstrated that aggression may not always be associated with negative social outcomes, and under some circumstances, may confer benefits (see Hawley, Little, & Rodkin, 2007, for review). For example, for children in high-aggression classrooms, aggressive behavior is not linked to peer rejection (Stormshak et al., 1999). Children identified by teachers as popular and aggressive are seen as 'cool' by peers in aggressive social networks (Rodkin, Farmer, Pearl, & Van Acker, 2006), and work with lower-income adolescents has indicated that aggressive-disruptive behavior is associated positively with perceived popularity (Luthar & McMahon, 1996).

Although some peer groups may approve of aggression, teachers likely will not. Indeed, engaging in physical and verbal aggression is a common reason students are suspended (Mendez & Knoff, 2003). This disparity presents youth with a challenging social problem: identifying and enacting behaviors approved of by multiple groups who may disagree about the efficacy of possible strategies. Previous work has suggested that such inter-judge discrepancies extend to behaviors other than aggression. For example, a recent study asked children in grades 2 and 5, their parents, and teachers to list behaviors associated with social competence. Although there was a significant overlap among the three groups, adults generated many behaviors not identified by children, such as having good manners (Warnes, Sheridan, Geske, & Warnes, 2005).

Goals of Current Work

The current study sought to extend existing work by examining differences in judges' evaluations of the effectiveness of youth behavior, focusing on judge-specific discrepancies in perceptions of youth responses to provocation by peers. Our investigation aimed to identify the types of strategies judged to be effective in youths' social environments. This focus on behavioral efficacy is consistent with many definitions of social competence (see Rose-Krasnor, 1997). Several previous studies have assessed youth evaluations with respect to a particular goal by asking whether the strategy would result in a desired instrumental or relational outcome (e.g., Crick & Dodge, 1996; see also Hughes, Meehan, & Cavell, 2004). Goals are linked to behavioral strategies (e.g., Rose & Asher, 1999) and are likely to be a source of variability in ratings of effectiveness. For example, youth for whom dominance or revenge are central goals may be more likely to rate physical aggression as effective (see Lochman, Wayland, & White, 1993) whereas youth who prioritize maintaining the relationship

with the other person may prefer assertive solutions (Chung & Asher, 1996). Although goals and outcome expectations are important for understanding individual differences in evaluations, the aim of the current work was not to elucidate the criteria underlying judgments of effectiveness. Rather, the objective was to understand how teachers and peers, on average, perceive the effectiveness of various strategies children use when confronted with peer provocation.

Studying how youths' responses to key situations—such as peer provocation—are received by their peers and teachers may aid in the development of more efficacious social-skills interventions. Examining inter-judge discrepancies could provide information about why some behaviors may be resistant to change. If a valued group does not approve of a particular strategy, children may be reluctant to engage in that behavior. As noted by Ladd (1984), teaching children to use strategies not valued by peers may actually harm their social well-being. Similarly, training behaviors frowned upon by teachers may result in negative outcomes in the academic context. Furthermore, youth are likely to choose strategies viewed positively by the evaluators whose opinions they value most, such as peers, even if their actions cause problems with other judges, such as teachers. In these cases, clinicians may need to work with youth to select behaviors that are more acceptable across groups or determine whose judgment will be the most consequential in which situations.

We established a framework for the assessment of how youth responses to peer provocation are perceived in the social environment during an earlier investigation, in which we asked youth what they would say or do when confronted with physical, verbal, and relational aggression by a peer (Dirks, Treats, & Weersing, 2007b). This investigation revealed that youths' responses could be coded reliably into eight categories: physical aggression, verbal aggression, ending relationship with the aggressor, damaging the aggressor's relationships with others, doing nothing, seeking an explanation, telling an adult, and stating that the provocation crossed limits (i.e., verbal responses that communicate directly that the aggressor's actions are unacceptable). Furthermore, a significant number of responses could be coded into more than one category. In particular, a hybrid category of 'hostile assertiveness' emerged, characterized by seeking an explanation or stating that the provocation crossed limits in conjunction with physical or verbal aggression.

Previous work has treated such hostilely assertive responses as aggressive. For example, two recent studies had children state how they would respond to a range of peer-provocation scenarios (Hughes et al., 2004; Peets, Hodges, Kikas, & Salmivalli, 2007). In both investigations, strategies combining multiple response types were coded as the most aggressive response present (e.g., responses including physical aggression and a prosocial strategy were coded only as physical aggression). In adopting this coding scheme, researchers are assuming that the presence of a more assertive response such as seeking an explanation will not mitigate the problems associated with aggressive behaviors. However, people in the social environment may view such responses as more effective. Providing peers with an opportunity to explain their actions, even if it is done in a hostile manner, may yield better social outcomes than verbal aggression by itself. In this case, treating these hybrid responses as aggressive will underestimate children's social effectiveness. It may also be of clinical utility to examine whether assertive strategies are viewed as significantly more competent than responses that marry such approaches with hostility. For example, asking 'Why did you do that?' instead of the more aggressive 'What's your problem?' will confer few advantages socially if the target is not sensitive to the difference.

Summary and Hypotheses

In summary, this study addressed two central questions. Firstly, how did peers and teachers differ in their judgments of the effectiveness of youth responses to peer provocation? Secondly, how did the addition of physical or verbal aggression to other types of responses affect teacher and peer evaluations of these strategies? These data will contribute to a more nuanced understanding of how specific behavioral strategies enacted in the context of a highly relevant class of social situations are evaluated by important groups in youths' social environments.

Based on previous studies suggesting that at least some peers may approve of aggression (e.g., Luthar & McMahon, 1996; Rodkin et al., 2006), we expected that youth would rate physically, verbally, and relationally aggressive strategies as more effective than would teachers. In contrast, earlier work has suggested that teachers may evaluate telling an adult more positively than youth. For example, two studies with children aged 8–11 years indicated that some children did not identify seeking adult help as a strategy they would use when managing peer conflict (Newman, Murray, & Lussier, 2001), particularly when faced with relational aggression (Goldstein, Tisak, Persson, & Boxer, 2006). Furthermore, work with youth aged 12–14 years suggested that these early adolescents are unlikely to go to a teacher when dealing with an aggressive peer (Dirks et al., 2007b).

In addition to examining differences in peer and teacher evaluations of youth responses, we investigated the relative effectiveness of different strategies within each judge group. Children may find aggression more acceptable than adults but may not think it is the most effective strategy overall. Work examining children's responses to problematic situations with peers has suggested that they are more likely to use other types of responses such as assertiveness (e.g., Chung & Asher, 1996; Dirks et al., 2007b). For this reason, we hypothesized that within the youth group, seeking an explanation and stating that the provocation crossed limits would be seen as more effective strategies than physical, verbal, and relational aggression.

We also explored sex differences in youth evaluations of the behavioral strategies as well as patterns of ratings within the youth group. Although the literature concerning sex differences in relationally and verbally aggressive behavior is mixed, numerous investigations have shown boys to be more physically aggressive than girls (see Archer, 2004; Tapper & Boulton, 2004). For this reason, we expected (1) that boys would evaluate physical aggression as more effective than girls and (2) that boys may be more likely to see physical aggression as effective compared with other response options. In contrast, girls have been identified as more likely to engage in prosocial and assertive responding (see Rose & Rudolph, 2006), which led us to predict (1) that girls would rate seeking an explanation and stating that the provocation crossed limits as more effective than boys and (2) that girls may be more likely to rate strategies such as seeking an explanation and stating that the provocation crossed limits as effective relative to other strategies.

The second primary goal of this study was to examine evaluations of hostilely assertive responses combining aggressive and assertive strategies. In general, we hypothesized that responses involving seeking an explanation or stating that the provocation crossed limits would be rated as more effective than responses combining these strategies with physical or verbal aggression, and that these hybrid strategies, in turn, would receive higher ratings than 'pure' aggression. A significant interaction between judge and type of response was also expected, such that the presence of verbal or physical aggression in a hybrid response would lower teachers' evaluations more than

those of youth. In addition, we expected sex differences to emerge within the youth group. Given the aforementioned evidence that boys are more physically aggressive than girls, it was expected that the presence of physical aggression would affect girls' judgments of effectiveness more than boys'.

Methods

Participants

Participants in this study were recruited from youth attending the sixth, seventh, and eighth grades at two middle schools in the northeastern USA, as well as their teachers. Both schools were located in the same community, which is predominantly lower income (per capita annual income = $18 404; Connecticut State Department of Education, 2006, based on data from 2001). According to school records, 62 percent of students qualified for free or reduced-cost lunch. Based on class lists, 1567 students were available to participate. Active consent was obtained from all participants. For students who did not return a written consent form, parents were contacted by phone to give consent. Two hundred and twenty-six parents were not reachable (14 percent), and 59 declined to participate (4 percent). Two hundred and twenty-five children were not available on the day of testing (14 percent). The measure used in this study was part of a large battery, and 394 students (25 percent) were not able to complete it. Thus, the final sample consisted of 663 students (Mean age = 12.63, *SD* = .96, range = 10.83–15.25 years of age), or 42 percent of available participants. Of these students, 15 percent were non-Hispanic White, 12 percent were African-American, 53 percent were Hispanic, and 19 percent were another ethnicity; 30 percent were in grade 6, 35 percent were in grade 7, and 35 percent were in grade 8; and 48 percent were male.

Questionnaires were given to 57 teachers; 49 (86 percent) completed the measure. Of the teachers, 74 percent were female, and 35 percent taught grade 6, 39 percent taught grade 7, and 18 percent taught grade 8. Three teachers (6 percent) taught more than one of those grades, and one teacher did not identify the grade taught.

Measurement of Evaluations of the Effectiveness of Youth Responses to Peer Provocation

All teachers and youth completed a questionnaire that consisted of the six vignettes describing physical, verbal, or relational provocation by a peer. The situations were chosen because (1) other work identified them as relevant for lower-income youth (Farrell et al., 1998) or (2) they have appeared frequently in vignette-based measures designed to assess social information-processing patterns (e.g., Crick, 1995; Crick & Dodge, 1996). In each story, the aggressor was described as a 'good friend' of the protagonist. Studies have shown that children's social information-processing patterns vary as a function of the aggressor's identity (e.g., Dirks et al., 2007b; Peets et al., 2007). We chose to identify the aggressor as a good friend because other researchers have noted that managing conflict in a friendship is an important task for children and adolescents (see Rose & Asher, 1999). For youth, the sex of both the characters was matched to the participant's sex. Teachers were assigned randomly questionnaires that identified the characters as either both male or both female. This randomization was undertaken to ensure that a roughly equal number of questionnaires about boys and girls were completed. Fifty-five percent of teachers completed questionnaires describing female characters.

Each vignette was paired with the same 14 responses (see Appendix), and the order of responses was randomized. The responses included one strategy from each of the eight coding categories. Two examples of stating that the provocation crossed limits were presented: saying 'I thought we were friends' and saying 'That's a really mean thing to do/say'. The latter strategy was included for comparison with the verbally aggressive strategy 'You are really mean', a contrast that allowed us to assess if judges are sensitive to the difference between identifying the person or the action as problematic.

The remaining five responses represented combinations of strategies including seeking an explanation with verbal aggression, stating that the provocation crossed limits with verbal aggression, and physical aggression with seeking an explanation. In an earlier study (Dirks et al., 2007b), 7 percent of youth responses to the peer-provocation vignettes were coded as both seeking an explanation and verbal aggression, and 9 percent were coded as both stating that the provocation crossed limits and verbal aggression. These combined responses occurred more frequently than verbal aggression alone or combined with any other type of strategy (4 percent), as well as ending relationship with the aggressor (5 percent), damaging the aggressor's relationships with others (3 percent), and doing nothing (6 percent). In the current measure, most of the cross-coded responses were designed to incorporate elements of the 'pure' responses (see Table 1). For example, 'You're really mean. Why did you do/say that?' combined the strategies coded as verbal aggression and seeking an explanation. Standardizing responses in this way helped to ensure that differences in evaluations were

Table 1. Responses and Coding Categories

Category	Response
Physical aggression	Shove him/her
Verbal aggression	'You're really mean'.
Ending relationship with aggressor	Never speak to him/her again
Damaging aggressor's relationships with others	Tell all of his/her other friends what the boy/girl did and get them on his/her side
Doing nothing	Not say or do anything
Seeking an explanation	'Why did you do/say that?'
Telling an adult	Tell the teacher
Stating that the provocation crossed limits	'I thought we were friends' 'That's a really mean thing to do/say'.
Physical aggression + seeking an explanation	Say 'Why did you do/say that?' and shove him or her
Verbal aggression + seeking an explanation	'What's your problem?' 'You're really mean. Why did you do/say that?'
Verbal aggression + stating that the provocation crossed limits	'You're really mean. I thought we were friends'.
Verbal aggression + seeking an explanation + stating that the provocation crossed limits	Say 'Why would you be so mean? I thought we were friends'.

due to the incorporation of an additional strategy rather than to differences in the nature of the response. Other combined responses were included because they were often generated by youth (e.g., asking 'What's your problem?').

Participants were asked to rate how well each response would 'work to solve the problem' on a five-point scale ranging from 1 (*would not work at all*) to 5 (*would work really well*).

Procedure

Measures were administered as part of a larger battery of baseline youth-report assessments designed to evaluate a curriculum developed to promote socioemotional learning. All measures were administered before students began the intervention. As it was not a focus of the primary study, the measure assessing youth's evaluations of responses to peer provocation was included at the end of the battery, which likely contributed to completion rates. It is expected that fatigue or time constraints will cause a number of participants to 'drop out' during a measurement battery. Rates of completion were similar for questionnaires administered immediately prior to this measure. In one of the two schools, participants completed a second measure related to responses to provocation. In this school, the order of presentation of the measures was counterbalanced. Thus, evaluations of responses to provocation were completed alone, first, or second.

Results

Representativeness of Sample

Among youth who received parental consent and were present on the day of testing, 37 percent did not complete the measure assessing evaluations of responses to provocation. Chi-square analyses were completed to determine if age, sex, and ethnicity were associated with completion of the measure. These analyses revealed that ethnicity, which was coded as non-Hispanic White, African-American, Hispanic, or other, was associated significantly with completion, $\chi^2(3) = 13.69$, $p < .01$. The pattern of proportions suggested that non-Hispanic White youth (72 percent complete) and youth of other ethnicities (70 percent complete) were more likely to complete the measure than Hispanic or African-American youth (59 and 63 percent complete, respectively). There was also a significant relationship between sex and completion, $\chi^2(1) = 6.92$, $p < .01$. Females (67 percent complete) were more likely to finish the questionnaire than males (59 percent complete).

Comparing Youth and Teacher Ratings of the Effectiveness of Responses

Prior to conducting analyses, 25 participants (4 percent) who had circled only one type of response or made patterned responses on the questionnaire were identified and removed from subsequent analyses. The first set of analyses examined the association between judge identity (i.e., teacher or youth) and evaluations of responses coded into each of the eight coding categories, plus a ninth strategy coded as stating that the provocation crossed limits. Children's mean evaluations of effectiveness for the nine strategies varied from 2.00 to 2.60 out of 5, a range markedly more restricted than that observed in teachers' evaluations (1.18 to 3.27). Inspection of the data revealed that most of the variables were significantly skewed. Square-root transformations resulted in sufficiently normal distributions and were applied to each variable in these analyses.

To begin, a repeated-measures ANOVA was conducted with one within-subjects factor (Response Type) and one between-subjects factor (Judge Identity). Sex of rater, grade (attending if student and teaching if teacher), and version of questionnaire (entered as two dummy-coded variables comparing questionnaires completed first or second to questionnaires completed alone) were included as covariates in this and all subsequent analyses, including *post hoc* examinations of significant interactions. Due to violations of the sphericity assumption, the more conservative Greenhouse–Geisser adjustment to degrees of freedom is reported for all tests involving the within-subjects variable and is denoted with the subscript *g-g*. The primary predictor of interest was the interaction between Response Type and Judge Identity.

The effectiveness ratings as a function of both Judge Identity and Response Type are depicted in Figure 1. The omnibus analysis revealed a significant interaction between Response Type and Judge Identity, $F_{g-g}(4.65, 3018.96) = 18.27$, $p < .01$, $\eta^2_p = .03$, as well as a significant main effect of Response Type, $F_{g-g}(4.65, 3018.96) = 43.03$, $p < .01$, $\eta^2_p = .06$. The interaction between Response Type and sex of rater was not significant. Among the covariates there was a significant interaction between Response Type and grade, $F_{g-g}(9.29, 3018.96) = 2.68$, $p < .05$, $\eta^2_p = .01$, as well as Response Type and version, $F_{g-g}(4.65, 3018.96) = 4.20$, $p < .01$, $\eta^2_p = .01$. There was no significant main effect of Judge Identity or of any of the covariates.
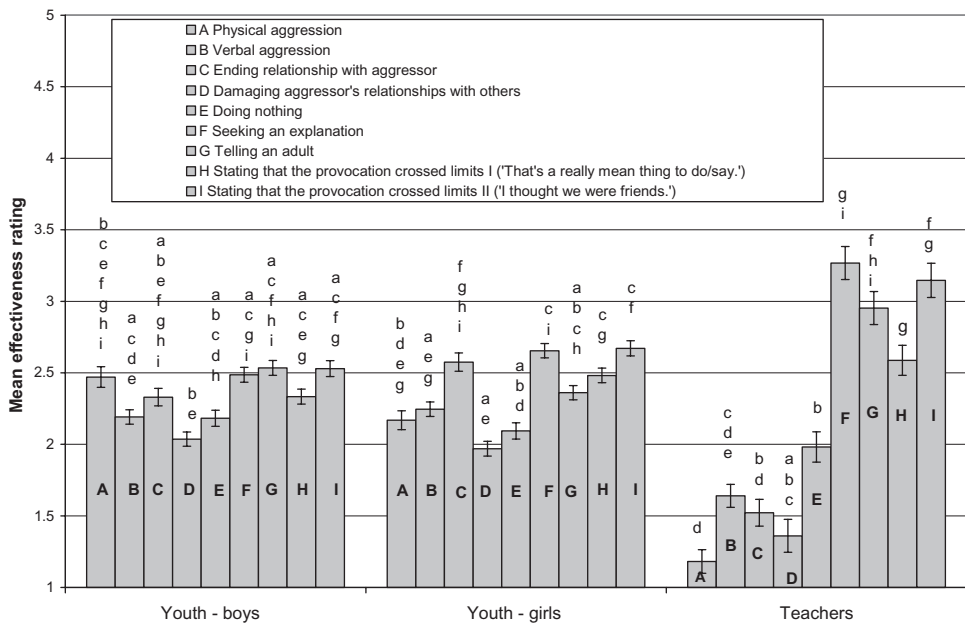
Figure 1. Evaluations of Effectiveness of Youth Responses to Peer Provocation as a Function of Response Type and Judge Identity.
Note: Error bars reflect the standard error of the mean for a given behavioral strategy. Analyses were conducted on square-root transformed variables, but bars are presented in original metric for ease of interpretation. Within each judge group, letters above a bar indicate that the mean rating of that category does not differ significantly from the mean rating of the bar bearing that letter, $p > .001$.

We decomposed the interaction between Response Type and Judge Identity in two ways. Firstly, to address the question of how peers and teachers differ in their ratings of youth responses, simple-effects analyses were conducted to compare teacher and youth ratings of each of the nine responses. Nine tests were completed; thus, alpha was set at .006 using a Bonferroni-correction procedure. These analyses indicated that youth evaluated the following types of responses as more effective than did teachers: physical aggression, $F(1, 3813.91) = 37.96$, $p < .006$, $d = 1.04$; verbal aggression, $F(1, 3813.91) = 8.48$, $p < .006$, $d = 0.66$; ending relationship with the aggressor, $F(1, 3813.91) = 28.74$, $p < .006$, $d = 0.91$; and damaging the aggressor's relationships with others, $F(1, 3813.91) = 15.66$, $p < .006$, $d = 0.78$. In contrast, teachers evaluated seeking an explanation as more effective than did youth, $F(1, 3813.91) = 13.37$, $p < .006$, $d = .75$. Teachers and youth did not differ in their evaluations of responses coded as stating that the provocation crossed limits, doing nothing, and telling an adult. It was expected that teachers would evaluate telling an adult as more effective than would youth, and teachers' ratings were in fact higher than those of youth, $d = .47$. A significant effect of grade, however, may have reduced the power of the analyses to detect statistically significant differences between judge groups for this response type. *Post hoc* Scheffé tests revealed that in general, telling an adult was seen as a significantly less effective strategy for eighth graders than for both sixth graders, $p < .001$, and seventh graders, $p < .05$.

The previous analyses provided evidence for the judge specificity of evaluations of competence. However, even though youth gave higher evaluations to aggressive responses than teachers, they might still rate other types of responses as more effective than aggression. When unpacking interactions between a within-subjects and a between-subjects factor, Howell (2002) has suggested conducting separate repeated-measures ANOVAs for each level of the between-subjects factor. These two analyses revealed a significant effect of Response Type on evaluations for both youth, $F_{g-g}(4.59, 2810.51) = 27.85$, $p < .01$, $\eta^2_p = .04$, and teachers, $F_{g-g}(4.54, 172.36) = 56.68$, $p < .01$, $\eta^2_p = .60$. Within the youth group, there was a significant interaction between sex of rater and Response Type, $F_{g-g}(4.59, 2810.51) = 7.65$, $p < .01$, $\eta^2_p = .01$; grade and Response Type, $F_{g-g}(9.19, 2810.51) = 3.67$, $p < .01$, $\eta^2_p = .01$; and version of questionnaire and Response Type, $F_{g-g}(4.59, 2810.51) = 4.09$, $p < .01$, $\eta^2_p = .01$. Within the teacher group, a significant interaction between grade and Response Type existed, $F_{g-g}(9.07, 172.36) = 2.03$, $p < .01$, $\eta^2_p = .10$.

To assess the pattern of evaluations within each judge group, repeated-measures ANOVAs were used to conduct all pair-wise comparisons of the covariate-adjusted means. Given the significant interaction between Response Type and sex of rater in the youth group, as well as our *a priori* interest in sex differences in response evaluation, analyses were conducted separately for boys and girls. Alpha was adjusted for the 35 comparisons made within each group and set at .001. Results of these analyses are presented in Figure 1. The pattern of findings indicated that teachers and girls rated seeking an explanation or stating that the provocation crossed limits ('I thought we were friends') as more effective than nearly all other responses. Boys, but not girls, viewed physical aggression to be as effective as these strategies. The ratings that boys and girls gave to the more assertive strategies were similar to their judgments of ending a relationship with the aggressor. Teachers, boys, and girls evaluated saying 'That was a really mean thing to do/say' (stating that the provocation crossed limits) as more effective than stating 'You are really mean' (verbal aggression), suggesting that both groups are sensitive to the difference between stating that there is a problem with the action or with the person.

Finally, to understand further the association between evaluations and demographic characteristics, we conducted simple-effects analyses to unpack the significant sex $\times$ Response Type interaction among the youth as well as the significant grade $\times$ Response Type interaction for youth and teachers. These analyses (alpha set at .006 using a Bonferroni-correction procedure) revealed that within the youth group, boys evaluated physical aggression as significantly more effective than did girls, $F(1, 3409.73) = 16.20$, $p < .006$, whereas girls evaluated ending relationship with the aggressor as significantly more effective than did boys, $F(1, 3409.73) = 7.60$, $p < .006$. Furthermore, grade was associated significantly with evaluations of telling an adult for youth, $F(2, 3409.73) = 9.88$, $p < .006$, and teachers, $F(2, 179.29) = 8.68$, $p < .006$. Comparison of covariate-adjusted means (alpha set at .017 using a Bonferroni-correction procedure) indicated that among youth, eighth graders evaluated this response as significantly less effective than did sixth graders, $F(1, 396) = 16.28$, $p < .01$. Eighth-grade teachers evaluated telling an adult as less effective than did both seventh-grade teachers, $F(1, 24) = 11.47$, $p < .01$, and sixth-grade teachers, $F(1, 21) = 11.47$, $p < .01$.

Evaluating Youth and Teacher Ratings of Responses Involving 'Hostile Assertiveness'

The second goal of this investigation was to examine the effect of combining aggressive and assertive responses on judges' evaluations. To address this question, a series of three repeated-measures ANOVAs was conducted. In each analysis, judges' mean evaluation of responses in the categories of interest served as the dependent variable. A square-root transformation was applied to each of these variables to reduce the skewness of the distributions. Separate analyses were conducted for each of the following three combinations of strategies: seeking an explanation and verbal aggression, stating that the provocation crossed limits and verbal aggression, and seeking an explanation and physical aggression. Each analysis consisted of one within-subjects factor, Response Type; one between-subjects factor, Judge Identity; and the interaction between these two factors. Sex of rater, grade (attending if student and teaching if teacher), and version of questionnaire were entered as covariates in these and all subsequent factorial analyses. The central questions addressed by these analyses focused on the main effect of Response Type and the Response Type $\times$ Judge Identity interaction. Each predictor was assessed three times; thus, alpha was set at .017.

The analysis examining the combination of verbal aggression and seeking an explanation revealed a main effect of Response Type, $F_{g\text{-}g}(2.44, 1587.56) = 67.17$, $p < .017$, $\eta^2_p = .10$, and a significant interaction between Response Type and Judge Identity, $F_{g\text{-}g}(2.44, 1587.56) = 28.41$, $p < .017$, $\eta^2_p = .04$. Examination of responses involving verbal aggression and stating that the provocation crossed limits yielded an identical pattern: a main effect of Response Type, $F_{g\text{-}g}(2.84, 1846.51) = 81.04$, $p < .017$, $\eta^2_p = .11$, and an interaction between Response Type and Judge Identity, $F_{g\text{-}g}(2.84, 1846.51) = 23.50$, $p < .017$, $\eta^2_p = .04$. In the analysis examining the combination of physical aggression and seeking an explanation, there was a significant main effect of Response Type, $F_{g\text{-}g}(1.31, 859.47) = 72.75$, $p < .017$, $\eta^2_p = .10$, and a significant interaction between Response Type and Judge Identity, $F_{g\text{-}g}(1.31, 859.47) = 43.16$, $p < .017$, $\eta^2_p = .06$. This analysis also revealed a main effect of Judge Identity, $F(1, 656) = 14.53$, $p < .017$, $\eta^2_p = .02$, and a significant interaction between version of questionnaire and Response Type, $F_{g\text{-}g}(1.31, 859.47) = 9.05$, $p < .017$, $\eta^2_p = .01$.

To follow up the significant interactions between Response Type and Judge Identity, a series of separate repeated-measures ANOVAs was conducted within the samples of youth and teachers. All six analyses revealed a significant main effect of Response Type. Results are reported in Figures 2–4. For the youth group, the analysis examining physical aggression and seeking an explanation yielded a significant interaction between sex of rater and Response Type, $F_{g\text{-}g}(1.31, 809.80) = 14.15$, $p < .017$, $\eta^2_p = .02$, and a significant interaction between version of questionnaire and Response Type, $F_{g\text{-}g}(1.31, 809.80) = 8.73$, $p < .017$, $\eta^2_p = .01$. There was also a significant interaction between grade and Response Type, $F_{g\text{-}g}(4.82, 1473.55) = 3.17$, $p < .017$, $\eta^2_p = .01$.

The repeated-measures ANOVAs were unpacked further by conducting pair-wise comparisons of the covariate-adjusted means. Comparisons among the responses coded in a single category (e.g., seeking an explanation vs. verbal aggression) were completed in an earlier analysis; thus, only tests involving the combination responses were conducted. Family-wise error rate was controlled by dividing .05 by the number of remaining comparisons for each type of strategy (four for verbal aggression and seeking an explanation, three for verbal aggression and stating that the provocation crossed limits, and two for physical aggression and seeking an explanation). Given the significant interaction between sex and Response Type when strategies involved physical aggression and seeking an explanation, *post hoc* tests for these strategies were conducted separately for boys and girls. Results of these analyses are presented in Figures 2–4.
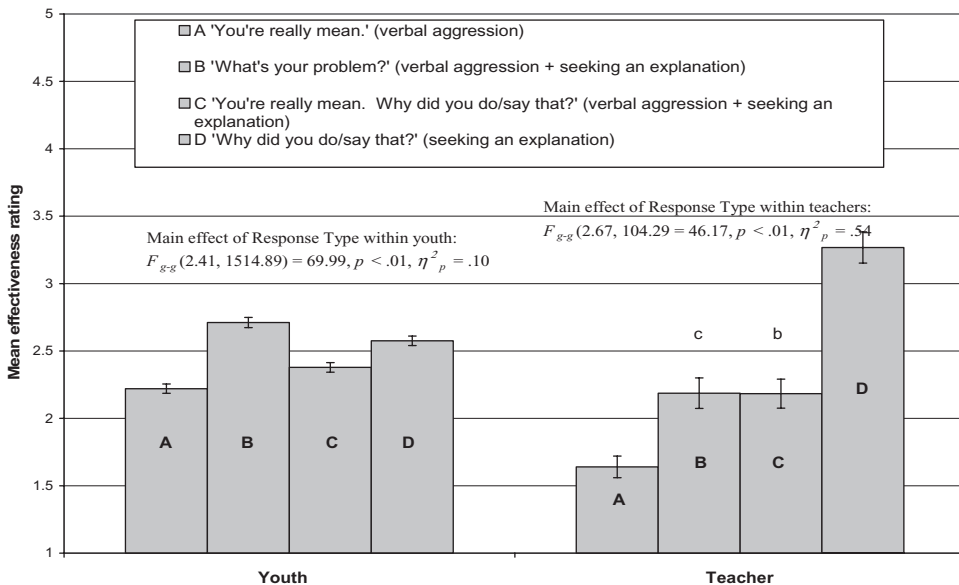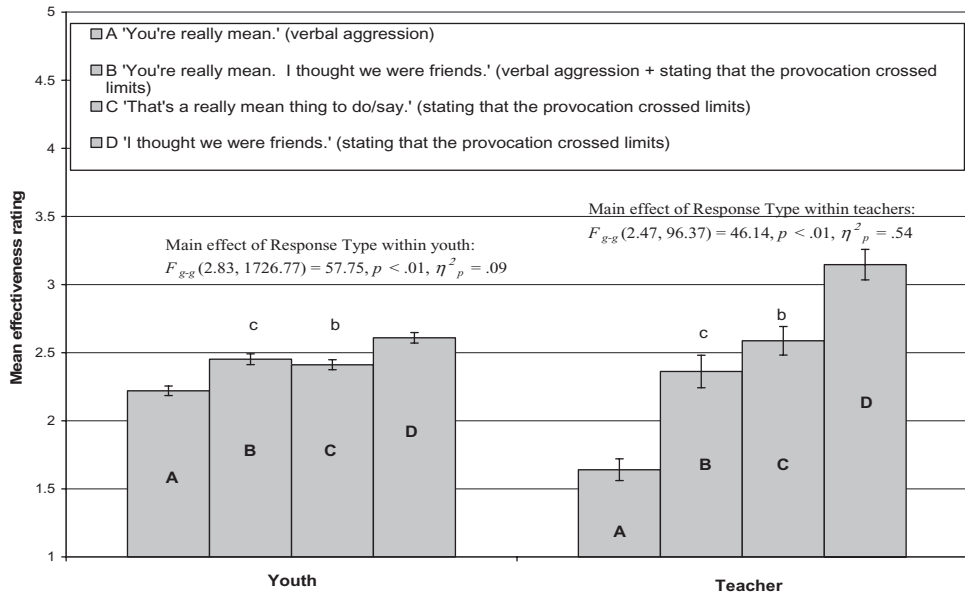

Figure 2. Evaluations of Effectiveness of Youth Responses Coded in Multiple Categories as a Function of Response Type and Judge Identity: Verbal Aggression and Seeking an Explanation.
Note: Error bars reflect the standard error of the mean for a given behavioral strategy. Analyses were conducted on square-root transformed variables, but bars are presented in original metric for ease of interpretation. Within each judge group, letters above a bar indicate that the mean rating of that category does not differ significantly from the mean rating of the bar bearing that letter, $p > .013$.

Figure 3. Evaluations of Effectiveness of Youth Responses Coded in Multiple Categories as a Function of Response Type and Judge Identity: Verbal Aggression and Stating that the Provocation Crossed Limits.
Note: Error bars reflect the standard error of the mean for a given behavioral strategy. Analyses were conducted on square-root transformed variables, but bars are presented in original metric for ease of interpretation. Within each judge group, letters above a bar indicate that the mean rating of the category does not differ significantly from the mean rating of the bar bearing that letter, $p > .017$.

Both groups evaluated responses combining verbal aggression and an assertive strategy as more effective than pure verbal aggression. Teachers evaluated responses involving only seeking an explanation to be more effective than the hostilely assertive responses. Among youth, on the other hand, asking 'What's your problem?' was judged more effective than the less aggressive question 'Why did you do/say that?' Both teachers and youth evaluated one of the limit-crossing strategies ('I thought we were friends') to be more effective than the response combining stating that the provocation crossed limits and verbal aggression. For girls and teachers, strategies combining physical aggression and seeking an explanation were rated as more effective than physical aggression only, and seeking an explanation was judged more effective than the combined response. In contrast, boys gave comparable ratings to all three responses.

Discussion

The first goal of this study was to examine differences between youth and teacher evaluations of the effectiveness of strategies generated by youth in response to peer provocation. As hypothesized, youth judged physical aggression to be more effective than did teachers. Previous studies have found that children who engage in physically aggressive behaviors may be perceived as popular (Luthar & McMahon, 1996) or 'cool' (Rodkin et al., 2006). Our work indicated that some youth also viewed
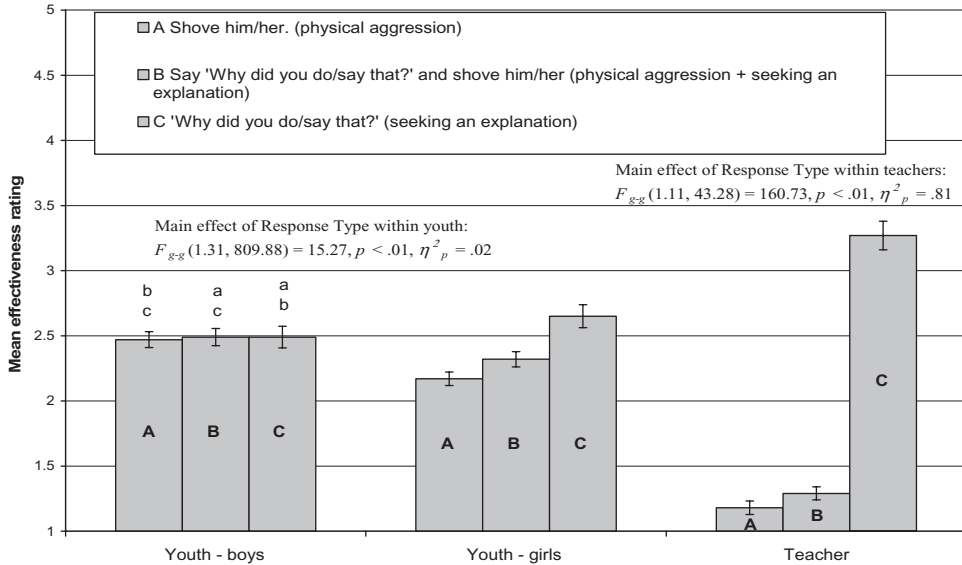
Figure 4. Evaluations of Effectiveness of Youth Responses Coded in Multiple Categories as a Function of Response Type and Judge Identity: Physical Aggression and Seeking an Explanation.

Note: Error bars reflect the standard error of the mean for a given behavioral strategy. Analyses were conducted on square-root transformed variables, but bars are presented in original metric for ease of interpretation. Within each judge group, letters above a bar indicate that the mean rating of the category does not differ significantly from the mean rating of the bar bearing that letter, $p > .025$.

physically aggressive behaviors as an effective way for their peers to manage conflict. Not only did teachers view this strategy as less effective than did youth, they also evaluated physical aggression as significantly less effective than nearly all other types of responses including both verbal aggression and ending one's relationship with the aggressor. Among youth, boys judged physical aggression to be more effective than did girls. In fact, boys viewed physical aggression to be as effective as all other strategies whereas girls judged physical aggression to be less effective than ending one's relationship with the aggressor, seeking an explanation, and stating that the provocation crossed limits. This sex difference is consistent with work that has shown boys to be more physically aggressive than girls (see Archer, 2004) and suggests that such behavior may be reinforced particularly within social networks comprising boys.

  Boys judged physically aggressive strategies to be more effective than damaging the aggressor's relationships with others, and in general, youth perceived this relationally aggressive strategy to be significantly less effective than nearly all other responses. Previous work has suggested that children believe that physical aggression is more 'wrong' and harmful than relational aggression (Murray-Close, Crick, & Galotti, 2006), although girls have reported that they would be hurt equally by relational and physical aggression (Galen & Underwood, 1997). Our findings indicated that on average, both male and female participants did not think that this type of relational aggression was especially effective.

Teachers did not differentiate between damaging the aggressor's relationship with others and ending one's relationship with the aggressor, a second strategy that may be construed as relationally aggressive (e.g., Delveaux & Daniels, 2000). In contrast, youth perceived ending one's relationship with the aggressor to be more effective than damaging the aggressor's relationships with others, and both boys and girls judged this strategy to be as effective as more assertive responses such as seeking an explanation. Differences in ratings of ending one's relationship with the aggressor and damaging the aggressor's relationships with others suggest that it is helpful to examine these two types of responses separately. The utility of this split is substantiated further by the finding that girls evaluated ending one's relationship with the aggressor to be more effective than did boys whereas girls and boys did not differ in their evaluations of damaging the aggressor's relationships with others. It may be that not all youth construe ending one's relationship with the aggressor as aggressive, particularly in the context of being provoked. In another study, a confirmatory factor analysis indicated that this strategy loaded strongly on two factors, aggression and avoidance/withdrawal, whereas damaging the aggressor's relationships with others loaded only on the aggression factor (Dirks, Treat, & Weersing, 2009). This pattern suggests that some children may perceive ending one's relationship with the aggressor as a method of avoiding the problem rather than confronting it directly.

Teachers and youth did not differ in their judgments of how well telling an adult would work to solve peer-provocation situations, an unexpected finding given that previous work has indicated that going to an adult is not a strategy early adolescents use often when managing peer provocation (Dirks et al., 2007b). Teachers did rate telling an adult as more effective than did youth ($d = .47$); however, this difference was not statistically significant. The power to detect this effect may have been reduced by a significant association between grade and both teacher and youth ratings. In general, both groups of judges perceived telling an adult to be less effective for eighth-grade students. Among children, this pattern may reflect age-related changes in perceptions of what constitutes effective social behavior. Finding that teachers' ratings vary as a function of the grade of the children in the vignettes suggests that the perceived competence of an action may be influenced by the actor's age, a key child-level characteristic. Grade was only associated with responses coded as telling an adult; however, the age range of the sample was relatively narrow. Age may have influenced ratings of additional categories had we included elementary and high school students.

Teachers did not rate any strategy as significantly more effective than telling an adult. Interestingly, neither did boys, whereas girls deemed seeking an explanation or stating 'I thought we were friends' more effective responses. Our results may reflect the tendency evident in the literature for girls to engage in more assertive behaviors (see Rose & Rudolph, 2006). It may also be that girls and boys expect different outcomes to result from telling the teacher. Newman et al. (2001) found that children who were less concerned about maintaining the friendship were more likely to seek help from the teacher. There is some evidence that maintaining relationships is more important to girls than to boys (Delveaux & Daniels, 2000), which may be, in part, why girls rate the assertive strategies as more effective than telling an adult. Although girls and boys differ in their perceptions of the relative effectiveness of seeking an explanation and stating the provocation crossed limits, the two groups did not differ significantly in their evaluations of these strategies. Girls did rate these strategies as more effective than did boys; however, the restricted range of evaluations within the youth group may have reduced the power of the analyses to detect these differences.

Among both teachers and youth, seeking an explanation and stating that the provocation crossed limits were highly rated responses. An earlier study revealed that these two types of actions are also the most commonly generated strategies in response to the six provocation vignettes (Dirks et al., 2007b). This work was conducted in a separate sample of ethnically diverse, lower-income youth, suggesting that many children are managing these provocation situations effectively, from the perspective of both peers and teachers.

The same study suggested that many youth give responses coded into multiple categories. In particular, a significant portion of responses combined seeking an explanation and verbal aggression (7 percent) or stating that the provocation crossed limits and verbal aggression (9 percent). Given the frequency with which such responses appear to be used, it is important to understand how they are received in children's social environments. Thus, the second goal of the current study was to examine how the addition of aggression to more assertive strategies affected youth and teacher judgments of effectiveness. Analyses indicated that for both teachers and youth, there was a significant main effect of Response Type for combinations of physical aggression and seeking an explanation, verbal aggression and seeking an explanation, and verbal aggression and stating that the provocation crossed limits. In general, responses combining physical or verbal aggression with an assertive response were evaluated as less effective than 'pure' assertion, and these combined responses were rated as more effective than 'pure' aggression. One exception was boys' ratings of strategies combining physical aggression and seeking an explanation. Unlike girls, whose pattern of evaluations matched that of the teachers, boys gave comparable ratings to physical aggression, physical aggression and seeking an explanation, and seeking an explanation. As noted earlier, boys evaluated physical aggression as quite effective relative to other strategies. Thus, from boys' perspective, adding an assertive element may not do much to improve the effectiveness of this response.

These findings have implications for both measurement and intervention models of youth social competence. Some assessments of youth social behavior place responses combining multiple strategies into the most aggressive category (e.g., Hughes et al., 2004; Peets et al., 2007). Our data indicated that this approach may result in loss of information, as responses combining aggression and another type of response may not be received in the same way as 'pure' aggression. Youth who utilize aggression in combination with assertion may be perceived by both peers and teachers as more socially effective than their peers who respond with only hostility. As such, it may be beneficial to help youth who interact aggressively to incorporate assertive elements, such as asking provocateurs why they behaved aggressively, into their responses to provocation. Even in the presence of aggression, inclusion of these strategies may enhance the perception of the efficacy of these responses in the social environment. Furthermore, it may help the child to obtain valuable information that could be useful in de-escalating the conflict.

As hypothesized, there was a significant interaction between presence of verbal or physical aggression and judge identity. In general, assertive responses were evaluated as more effective than responses combining aggression with these strategies. Overall, it appears that the negative impact of physical and verbal aggression on effectiveness ratings is far greater for teachers than for peers. This finding suggests that teaching youth not to engage in aggression may improve relationships with teachers more than peers. An alternate way to consider these data is that responses combining verbal aggression and assertion may, for some youth, strike a balance among different groups'

perceptions of what constitutes effective behavior. These responses are rated as significantly more effective than pure aggression by teachers and appear to have broad support among peers. In fact, youth rated the hostilely assertive response 'What's your problem?' as more effective than the less aggressive 'Why did you do/say that?'

In general, we observed greater differences in teacher evaluations of how well different types of responses would work to solve the problem, relative to youth. Repeated-measures ANOVAs conducted on each group's evaluations revealed that the effect size associated with Response Type for teachers was $\eta^2_p = .60$ compared with $\eta^2_p = .04$ for youth. These differences may be partially due to the markedly higher variability in youth's evaluations of each response. For example, the *SD* associated with youth's ratings of physical aggression was .40 vs. .20 for teachers. Large differences in *SD*s were observed for other response types including verbal aggression (.30 for youth and .21 for teachers) and ending relationship with the aggressor (.36 for youth and .25 for teachers). These differences suggest that for youth, factors besides type of response are contributing to evaluations. In contrast, for teachers, characteristics of the strategy may be the most important determinant of their ratings.

Undoubtedly, then, multiple factors contributed to the observed differences in peers' and teachers' evaluations, and it will be important for future work to unpack the criteria underlying each group's judgments. One likely source of variability is the goals that youth and teachers may have had in mind when evaluating behaviors. For example, some youth may have been fixated on revenge, which has been associated with use of aggressive strategies (e.g., Lochman et al., 1993), whereas teachers may have prioritized maintaining the relationship with the person or ending the conflict. Judgments of effectiveness may also have been influenced by the circumstances in which peers and teachers typically observe youth behavior. Generally, teachers interact with youth in classrooms, a context where aggressive behavior will be even less appropriate than the schoolyard or children's neighborhoods. Although the vignettes specify where the provocation occurred, teachers and peers may still bring their experiences in different contexts to bear on their evaluations. Finally, participants were asked to rate the effectiveness of behavior. In making these judgments, teachers may have been more influenced by moral beliefs about what is 'right' and 'wrong' in a situation than youth, who may have focused more exclusively on what is adaptive or functional.

Limitations

It is important to note several limitations of the current study. Firstly, a significant percentage of the available sample did not complete the measure and rates of completion were associated with variables of interest including ethnicity and sex. Although this attrition is troubling, available evidence suggests that youth with greater behavior problems may be less likely to complete the measure due to not wanting or being able to participate. At least one previous study has shown that youth high on distress and low on self-restraint (i.e., suppression of aggression, impulse control, consideration of others) were considerably more likely not to participate in a school-based data collection (Weinberger, Tublin, Ford, & Feldman, 1990). A second study demonstrated that aggressive youth left a longitudinal study at greater rates than their typically developing peers (Metropolitan Area Child Study Research Group, 2002). As the extant literature has indicated that aggressive youth are more likely to evaluate aggressive strategies positively (see Dodge, 1993), the inclusion of these youth might have magnified the differences between student and teacher ratings rather than attenuating our findings.

The fact that our sample consisted primarily of lower-income youth likely contributed to the relatively high rate of non-completion. Several factors associated with lower socioeconomic status (SES), such as increased behavioral difficulties (Tolan, Gorman-Smith, Huesmann, & Zelli, 1997) and attending a school with elevated rates of absenteeism, suspensions, and truancy (Peng & Lee, 1993), may elevate the likelihood of not completing research questionnaires. Perhaps due in part to these challenges, lower-SES youth remain an understudied group in psychological research despite being at greater risk for psychosocial difficulties (e.g., Costello, Compton, Keeler, & Angold, 2003). Specific to the questions considered in this investigation, lower-income youth may be more likely to function in social environments in which considerable disagreement exists between youth and adults about the appropriateness of certain social behaviors (see Luthar & Burack, 2000). Thus, it is especially critical to understand how these youth and their teachers evaluate social strategies.

A second potential limitation of this study is the restricted range of youths' ratings of behavioral strategies. There was markedly more variability among youth in their evaluations of each strategy as compared with teachers. For example, the *SD* of youth evaluations of physical aggression was .40 compared with .20 for teachers. The difference between mean evaluations of each strategy, however, was greater for teachers. Mean teacher ratings ranged from 1.18 to 3.27 whereas mean youth ratings ranged from 2.00 to 2.60. In other words, there was greater within-strategy variability for youth and greater between-strategy variability for teachers. On average, then, youth ratings were restricted to the lower end of the scale and were relatively similar across behavioral strategies, a pattern likely due to several factors. Firstly, these situations were rated as difficult by youth (Dirks, 2007). It may be that no response is seen as particularly effective in the context of very challenging situations. Secondly, the response set was highly standardized. Responses were designed so that identical phrases appeared across a number of strategies (e.g., 'You're really mean'), and an identical set of responses was presented across all vignettes. This approach was used to increase the internal validity of comparisons of evaluations of responses combining different strategies. However, the standardization of responses precluded our tailoring strategies for each scenario, and some youth may prefer responses that are well crafted to fit the provocation closely. This standardization may have affected teachers' evaluations less, as it may be easier for adults to generalize from the presented strategy to other, similar responses that provide a better fit to the situation. As the focus of the current study was on comparing evaluations, the internal validity of the comparisons was prioritized, although this may have contributed to fewer responses at the upper end of the scale from youth.

Summary and Conclusion

This study is among the first to provide explicit empirical evidence for the judge specificity of evaluations of the effectiveness of youth behavior. Previous studies have suggested that such differences might exist, particularly for judgments of physical aggression (e.g., Luthar & McMahon, 1996). Such work has established that physical aggression may be viewed positively among some youth but does not unpack the criteria underlying those judgments nor establish how others in the social environment might perceive those responses.

Our findings may have implications for measurement and intervention models of youth social competence. Theorists are recognizing increasingly that social competence is an evaluative construct that will vary as a function of judge (see Dirks et al.,

2007a). However, most studies that have assessed social competence from the perspective of multiple judges have combined informants' ratings, thus treating inter-rater discrepancy as error. The present data suggest that some of this variation in judges' perceptions reflects meaningful differences that shape youth's social behavior and impacts their social functioning with important people in their lives. Further understanding of differences in judges' perceptions of the efficacy of youth's social strategies could be used to inform assessment and intervention efforts. One common method of measuring youth social functioning is behavior rating scales, in which respondents rate how often the targeted child engages in behaviors prejudged to be problematic or acceptable. The current study suggests that the perceived skillfulness of these behaviors may vary in meaningful ways as a function of who is evaluating them. Thus, knowing how children behave does not reveal the whole story of their social functioning. It is also important to understand how those behaviors will be received by people in the social environment.

Developing measures that capture variability in the perceived efficacy of youths' social strategies may provide a more detailed picture of their social functioning, thereby serving as the foundation for more focused interventions. Social-skills training should target behaviors important for success in a given child's social environment (Sheridan, Hungelmann, & Maughan, 1999). Assessing which responses are problematic for which types of judges will pinpoint behaviors likely to have the greatest positive or negative impact on social functioning. Additionally, illuminating inter-judge discrepancies may help clinicians work with their young clients to develop responses to key social situations that are perceived as effective (or, at least, not grossly ineffective) by multiple relevant groups. Ultimately, then, this line of inquiry may inform researchers and practitioners aiming to improve youth's social functioning.

References

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

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

Bierman, K. L. (2004). *Peer rejection: Developmental processes and intervention strategies*. New York: Guilford.

Chung, T.-Y., & Asher, S. R. (1996). Children's goals and strategies in peer conflict situations. *Merrill-Palmer Quarterly*, *42*, 125–147.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development*, *59*, 815–829.

Cole, D. A., Gondoli, D. M., & Peeke, L. G. (1998). Structure and validity of parent and teacher perceptions of children's competence: A multitrait-multimethod-multigroup investigation. *Psychological Assessment*, *10*, 241–249.

Connecticut State Department of Education. (2006). *Strategic school profile 2005–2006: New Britain public schools*. Hartford: Connecticut Department of Education.

Connolly, J., & Doyle, A. (1981). Assessment of social competence in preschoolers: Teachers versus peers. *Developmental Psychology*, *17*, 454–462.

Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association*, *290*, 2023–2029.

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology*, *7*, 313–322.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, *67*, 993–1002.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment*, *16*, 330–334.

Delveaux, K. D., & Daniels, T. (2000). Children's social cognitions: Physically and relationally aggressive strategies and children's goals in peer conflict situations. *Merrill-Palmer Quarterly*, *46*, 672–692.

Dirks, M. A. (2007). *Situation- and judge-level factors in the conceptualization and measurement of youth social functioning: The case of peer provocation*. Unpublished doctoral dissertation.

Dirks, M. A., Treat, T. A., & Weersing, V. R. (2007a). Integrating theoretical, measurement, and intervention models of youth social competence. *Clinical Psychology Review*, *27*, 327–347.

Dirks, M. A., Treat, T. A., & Weersing, V. R. (2007b). The situation specificity of youth responses to peer provocation. *Journal of Clinical Child and Adolescent Psychology*, *36*, 621–628.

Dirks, M. A., Treat, T. A., & Weersing, V. R. (2009). The latent structure of youth responses to peer provocation. Manuscript submitted for publication.

Dodge, K. A. (1985). Facets of social interaction and the assessment of social competence in children. In B. A. Schneider, K. H. Rubin, & J. E. Ledingham (Eds.), *Children's peer relations: Issues in assessment and intervention* (pp. 3–22). New York: Springer-Verlag.

Dodge, K. A. (1993). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology*, *44*, 559–584.

Dodge, K. A., Laird, R., Lochman, J. E., Zelli, A., & the Conduct Problems Prevention Research Group. (2002). Multidimensional latent construct analysis of children's social information processing patterns: Correlations with aggressive behavior problems. *Psychological Assessment*, *14*, 60–73.

Farrell, A. D., Ampy, L. A., & Meyer, A. L. (1998). Identification and assessment of problematic interpersonal situations for urban adolescents. *Journal of Clinical Child Psychology*, *27*, 293–305.

Farrell, A. D., Sullivan, T. N., Kliewer, W., Allison, K. W., Erwin, E. H., Meyer, A. L., et al. (2006). Peer and school problems in the lives of urban adolescents: Frequency, difficulty, and relation to adjustment. *Journal of School Psychology*, *44*, 169–190.

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

Goldstein, S. E., Tisak, M. S., Persson, A. V., & Boxer, P. (2006). Children's evaluations of ambiguous provocation by relationally aggressive, physically aggressive and prosocial peers. *British Journal of Developmental Psychology*, *24*, 701–708.

Hawley, P., Little, T. D., & Rodkin, P. C. (Eds.). (2007). *Aggression and adaptation: The bright side to bad behavior*. Mahwah, NJ: Erlbaum.

Howell, D. C. (2002). *Statistical methods for psychology*. Pacific Grove, CA: Duxbury.

Hughes, J. N., Meehan, B. T., & Cavell, T. A. (2004). Development and validation of a gender-balanced measure of aggression-relevant social cognition. *Journal of Clinical Child and Adolescent Psychology*, *33*, 292–302.

Kochenderfer-Ladd, B. (2004). Peer victimization: The role of emotions in adaptive and maladaptive coping. *Social Development*, *13*, 329–349.

Ladd, G. W. (1984). Social skill training with children: Issues in research and practice. *Clinical Psychology Review*, *4*, 317–337.

Ladd, G. W. (2005). *Children's peer relations and social competence: A century of progress*. New Haven, CT: Yale University Press.

Langlois, J. H., & Stephan, C. (1977). The effects of physical attractiveness and ethnicity on children's behavioral attributions and peer preferences. *Child Development*, *48*, 1694–1698.

Lochman, J. E., Wayland, K. K., & White, K. J. (1993). Social goals: Relationship to adolescent adjustment and to social problem solving. *Journal of Abnormal Child Psychology*, *21*, 135–151.

Luthar, S. S., & Burack, J. A. (2000). Adolescent wellness: In the eye of the beholder? In D. Cicchetti, J. Rappaport, I. Sandler, & R. P. Weissberger (Eds.), *The promotion of wellness in children and adolescents* (pp. 29–57). Washington, DC: CWLA Press.

Luthar, S. S., & McMahon, T. J. (1996). Peer reputation among inner-city adolescents: Structure and correlates. *Journal of Research on Adolescence*, *6*, 581–603.

Mendez, L. M. R., & Knoff, H. M. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children*, *26*, 30–51.

Metropolitan Area Child Study Research Group. (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology*, *70*, 179–194.

Murray-Close, D., Crick, N. R., & Galotti, K. M. (2006). Children's moral reasoning regarding physical and relational aggression. *Social Development*, *15*, 345–372.

Newman, R. S., Murray, B., & Lussier, C. (2001). Confrontation with aggressive peers at school: Students' reluctance to seek help from the teacher. *Journal of Educational Psychology*, *93*, 398–410.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin*, *102*, 357–389.

Peets, K., Hodges, E. V. E., Kikas, E., & Salmivalli, C. (2007). Hostile attributions and behavioral strategies in children: Does relationship type matter? *Developmental Psychology*, *43*, 889–900.

Peng, S. S., & Lee, R. M. (1993). *Educational experience and needs of middle school students in poverty*. Washington, DC: National Center for Education Statistics, US Department of Education.

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*, *24*, 239–254.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2006). They're cool: Social status and peer group supports for aggressive boys and girls. *Social Development*, *15*, 175–204.

Rose, A. J., & Asher, S. R. (1999). Children's goals and strategies in response to conflicts within a friendship. *Developmental Psychology*, *35*, 69–79.

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of boys and girls. *Psychological Bulletin*, *132*, 98–131.

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development*, *6*, 111–135.

Rudolph, K., Hammen, C., Burge, D., Lindbergh, N., Hertzberg, D., & Daley, S. E. (2000). Toward an interpersonal life stress model of depression: The developmental context of stress generation. *Development and Psychopathology*, *12*, 215–234.

Schneider, B. H., & Byrne, B. M. (1989). Parents rating children's social behavior: How focused the lens? *Journal of Clinical Child Psychology*, *18*, 237–241.

Sheridan, S. M., Hungelmann, A., & Maughan, D. P. (1999). A contextualized framework for social skills assessment, intervention, and generalization. *School Psychology Review*, *28*, 84–103.

Slaby, R. G., & Guerra, N. G. (1988). Cognitive mediators of aggression in adolescent offenders: I. Assessment. *Developmental Psychology*, *24*, 580–588.

Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., Coie, J. D., & the Conduct Problems Prevention Research Group. (1999). The relation between behavior problems and peer preference in different classroom contexts. *Child Development*, *70*, 169–182.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Socially undesirable need not be incompetent: A response to Crick and Dodge. *Social Development*, *8*, 132–134.

Tapper, K., & Boulton, M. J. (2004). Sex differences in levels of physical, verbal, and indirect aggression amongst primary school children and their associations with beliefs about aggression. *Aggressive Behavior*, *30*, 123–145.

Tolan, P. H., Gorman-Smith, D., Huesmann, L. R., & Zelli, A. (1997). Assessing family process to explain risk for antisocial behavior and depression among urban youth. *Psychological Assessment*, *9*, 212–223.

Warnes, E. D., Sheridan, S. M., Geske, J., & Warnes, W. A. (2005). A contextual approach to the assessment of social skills: Identifying meaningful behaviors for social competence. *Psychology in the Schools*, *42*, 173–187.

Weinberger, D. A., Tublin, S. K., Ford, M. E., & Feldman, S. S. (1990). Preadolescents' social-emotional adjustment and selective attrition in family research. *Child Development*, *61*, 1374–1386.

Acknowledgments

Appendix

Sample Page of Questionnaire Assessing Evaluations of Youth Responses to Peer Provocation

Kayla has just started walking home after school is over. She can see other students and teachers leaving for the day as well. She sees her good friend walking quickly towards her. Her friend looks really angry. She walks right up to Kayla, and before Kayla can say anything, her friend shoves her.

Below is a list of things that Kayla might say or do in this situation.

How WELL would each of these things WORK TO SOLVE THE PROBLEM?

	WOULD NOT WORK AT ALL		WOULD WORK OKAY		WOULD WORK REALLY WELL
1. Say 'Why did you do that?'	1	2	3	4	5
2. Say 'What's your problem?'	1	2	3	4	5
3. Shove her	1	2	3	4	5
4. Say 'I thought we were friends.'	1	2	3	4	5
5. Tell all of her other friends what the girl did and get them on her side	1	2	3	4	5
6. Say 'Why did you do that?' and shove her	1	2	3	4	5
7. Never speak to her again	1	2	3	4	5
8. Say 'You're really mean. Why did you do that?'	1	2	3	4	5
9. Say 'That's a really mean thing to do.'	1	2	3	4	5
10. Not say or do anything	1	2	3	4	5
11. Say 'Why would you be so mean? I thought we were friends.'	1	2	3	4	5
12. Tell the teacher	1	2	3	4	5
13. Say 'You're really mean. I thought we were friends.'	1	2	3	4	5
14. Say 'You're really mean.'	1	2	3	4	5