

# Reducing Playground Bullying and Supporting Beliefs: An Experimental Trial of the *Steps to Respect* Program

Karin S. Frey

University of Washington and Committee for Children

Miriam K. Hirschstein, Jennie L. Snell,

Leihua Van Schoiack Edstrom,

Elizabeth P. MacKenzie, and Carole J. Broderick  
Committee for Children

Six schools were randomly assigned to a multilevel bullying intervention or a control condition. Children in Grades 3–6 ( $N = 1,023$ ) completed pre- and posttest surveys of behaviors and beliefs and were rated by teachers. Observers coded playground behavior of a random subsample ( $n = 544$ ). Hierarchical analyses of changes in playground behavior revealed declines in bullying and argumentative behavior among intervention-group children relative to control-group children, increases in agreeable interactions, and a trend toward reduced destructive bystander behavior. Those in the intervention group reported enhanced bystander responsibility, greater perceived adult responsiveness, and less acceptance of bullying/aggression than those in the control group. Self-reported aggression did not differ between the groups. Implications for future research on the development and prevention of bullying are discussed.

Bullying in school is a pervasive social problem in which children exploit power imbalances in order to dominate and harm others physically, socially, or emotionally (Olweus, 1993; Smith & Brain, 2000). Research from the literature on bully victims and general victimization shows that involvement in this aggressive process is associated with poor outcomes for those who bully (Andershed, Kerr, & Stattin, 2001; Connolly, Pepler, Craig, & Taradash, 2000; Nansel et al., 2001) as well as for those who are

victims of bullying (for a meta-analysis, see Hawker & Boulton, 2000) or bystanders to bullying (O'Connell, Pepler, & Craig, 1999). Thus, bullying potentially affects an entire school.

Olweus's (1993, 1999) seminal intervention research has emphasized the importance of viewing bullying within a multilevel social context, whereas developmental researchers have uncovered individual characteristics likely to contribute to the processes underlying bullying and victimization. Together, these streams of research suggest that possible risk factors include (a) lack of adult awareness and systemic supports to prevent bullying (Pepler, Craig, & O'Connell, 1999; Olweus, 1993), (b) destructive bystander behavior (O'Connell et al., 1999; Salmivalli, 1999), (c) student beliefs that support bullying (Owens, Slee, & Shute, 2000; Pellegrini, 2002), and (d) student social-emotional skill deficits (Kochenderfer & Ladd, 1997; Schwartz, 2000).

---

Karin S. Frey, Department of Educational Psychology, University of Washington, and Committee for Children, Seattle, Washington; Miriam K. Hirschstein, Jennie L. Snell, Leihua Van Schoiack Edstrom, Elizabeth P. MacKenzie, and Carole J. Broderick, Committee for Children, Seattle, Washington.

Elizabeth P. MacKenzie and Jennie L. Snell are now at the University of Washington in the Social Development Research Group and the Department of Psychology, respectively. Carole J. Broderick is now at the State of Washington Department of Social and Health Services, Division of Alcohol and Substance Use.

This research was supported by the Committee for Children. The views expressed herein are the authors' and have not been cleared by the grantors.

We thank the school personnel, the children, and their parents not only for taking on the challenges of implementing a prevention program but for consenting to participate in this study. We also thank the following individuals: study coordinators Catherine Keelan and Truc Nguyen; teacher consultant Virginia Blashill; John Hilgedick and Rose Kroeker for creating the Personal Digital Assistant program and computer interface; Doug Cooper for manuscript assistance; and Sue Nolen and Diane Jones for their thoughtful comments on an earlier version of the article. We are indebted to the team of skilled research assistants who carried out the playground observations, particularly Collin Revoir, who assisted both with coding and coder training. Finally, we thank Debra Pepler and Wendy Craig for generously sharing their vast knowledge and collection of playground videotapes.

Correspondence concerning this article should be addressed to Karin S. Frey, Department of Educational Psychology, Box 3600, University of Washington, Seattle, WA 98195. E-mail: kfrey@cfchildren.org

## Adult and Systemic Factors

Although teachers perceive themselves as intervening often against bullying, observational research shows teachers intercede in only 15% to 18% of classroom bullying episodes (Craig, Pepler, & Atlas, 2000). Compounding the problem is that many students do not report being bullied to school staff, perhaps because of a perception that reporting rarely leads to effective intervention and entails risks of retaliation (Hoover, Oliver, & Hazler, 1992). Many leaders in the field argue that changing the dynamics of bullying requires increasing adult awareness and intervention, developing clear school policies, and coordinating procedures to track and respond to bullying reports (e.g., Olweus, 1993; Pepler, Craig, & O'Connell, 1999; Smith & Sharp, 1994).

## Bystander Behavior

Bystanders to bullying events may contribute to the problem by providing attention and assistance to those who bully. Live observations showed bystanders involved in more than 80% of bullying episodes and generally reinforcing the aggression. Peers inter-

vened rarely, but when they did, bullying tended to stop quickly (Hawkins, Pepler, & Craig, 2001). This finding suggests that increasing bystanders' socially responsible behavior, and the skills and beliefs that support its execution, may help reduce school bullying.

### Student Beliefs and Skills

There is a strong theoretical case for addressing the skills and beliefs of students in order to reduce bullying. Crick and Dodge's (1994) model of social behavior posits that behavior follows from the interaction of (a) beliefs and goals with (b) cognitive-behavioral and affective skills. Thus, effective student-level interventions would target both arenas.

#### *Beliefs About Bullying*

The aggression literature suggests that aggressive behavior and normative beliefs have a transactional relationship (Huesmann & Guerra, 1997; Ladd & Troop-Gordon, 2003). Although this relationship has not been tested in the bullying literature, descriptive work suggests a possible link between lack of empathy for victims and perpetrating and/or watching bullying (Endresen & Olweus, 2001; Pellegrini & Long, 2002). The belief that bullying is just fun and games or is justified if the targeted child is somehow "weird" may enable children to rationalize their participation as aggressors or spectators (Hoover et al., 1992; Owens et al., 2000; Rigby & Slee, 1991). Moreover, the belief that adults rarely intervene may lead children to conclude that one can bully with impunity.

#### *Social-Emotional Skills*

Bystanders may feel confused about the right course of action to take in the face of bullying (Craig et al., 2000), especially in a context of multiple, nonresponsive onlookers (Darley & Latane, 1968). Attempts to foster socially responsible behavior may include (a) providing clear guidelines for responding and reporting, (b) teaching strategies to cope with distressing emotions (Eisenberg & Fabes, 1998; Snell, MacKenzie, & Frey, 2002), and (c) practicing assertive, emotionally regulated responses to bullying (e.g., rehearsing "Stop. That is bullying.>").

Victimized students might also benefit from instruction in self-regulatory techniques and "scripts" to foster the calm, assertive responses associated with subsequent reductions in harassment (Schwartz, Dodge, & Coie, 1993). Moreover, teaching skills may reduce social-emotional deficits, a risk factor for victimization, and build possible protective factors such as agreeableness and friendship skills (Boulton, Trueman, Chau, Whitehand, & Amatya, 1999; Kochenderfer & Ladd, 1997; Perry, Hodges, & Egan, 2001). It is plausible therefore that skill instruction aimed at bullying reduction could increase general interpersonal skills.

The suitability of social-emotional skill instruction for children who bully has been debated (Sutton, Smith, & Swettenham, 1999), because some of these students demonstrate high levels of social intelligence and facility at manipulating others (Kaukiainen et al., 2002). Links between anger and bullying, however (Espelage, Bosworth, & Simon, 2001), suggest that training in emotion regulation and assertiveness (Salmivalli, 1999) may help reduce ag-

gression in some children, particularly those who are both aggressive and victimized (Schwartz, 2000).

### School-Based Outcome Research

Some large-scale evaluations of multilevel bullying prevention efforts have yielded promising results (for reviews, see Rigby, 2002; Smith & Ananiadou, 2003), such as substantial reductions in student reports of bullying (Olweus, 1993; Whitney, Rivers, Smith, & Sharp, 1994), improved attitudes toward school (Olweus, 1993, 1999; Whitney et al., 1994), and increased willingness to seek help when bullied (Whitney et al., 1994). Others have found quite modest improvements (Stevens, Van Oost & de Bourdeaudhuij, 2000) or none at all (Roland, 2000). Especially noteworthy are large-scale investigations that provide information about school implementation effects (Olweus, 1999; Roland, 2000; Stevens et al., 2000), albeit at the cost of more in-depth measurement efforts.

### The Current Study

In the current study, we took a complementary approach. Our multi-informant strategy included microanalytic observations of bullying, bystander, and adult behavior; teacher ratings; and self-reports of beliefs and behavior. Using a randomized control design, we evaluated the effects of a universal, school-based program, *Steps to Respect: A Bullying Prevention Program* (Committee for Children, 2001). Program components focus on (a) addressing adult and systemic factors through school policy development and staff training and (b) promoting prosocial beliefs and social-emotional learning through a classroom curriculum.

#### *Primary Study Goals*

The three primary study goals were to examine intervention effects on bullying and bystander behavior on playgrounds, on children's bullying-related beliefs, and on social-emotional skills. A secondary goal concerned examining program effectiveness as a function of grade, gender, and baseline levels of bullying.

*Goal 1: Reducing bullying and destructive bystander behaviors.* This goal was evaluated with playground observations and student self-reports. We predicted that the intervention group would show relative decreases in observed playground bullying compared with the control group. We also predicted relative decreases in bystander encouragement of bullying.

To ensure that our coding measure enabled sufficient discrimination of bullying versus nonbullying aggression, we measured both types of behavior. However, we had no a priori hypotheses regarding nonbullying aggression.

Although we expected a decline in self-reported bullying and victimization after program participation, we thought increased knowledge about what constitutes bullying might initially inflate behavioral reports (Rahey & Craig, 2002; Schafer, Werner, & Crick, 2002). We therefore predicted declines but did not know if they would occur after only 1 year of program implementation.

*Goal 2: Increasing prosocial beliefs related to bullying.* Self-report data addressed beliefs relevant to bullying, specifically, acceptance of bullying/aggression, spectator interest, and perceived responsibility to intervene in bullying. We predicted rela-

tive increases in prosocial beliefs among the intervention group. We also hypothesized that those in the intervention group, in contrast to those in the control group, would perceive adults as more knowledgeable and responsive in relation to bullying.

*Goal 3: Increasing social-emotional skills.* We predicted that the perceived difficulty of making a calm, strong response to bullying would decrease among intervention-group children relative to control-group children. Children's peer interaction skills were assessed via teacher report and observations of agreeable (Perry et al., 2001) and argumentative social interactions. We predicted that children in the intervention group, relative to those in the control group, would display greater skills on these measures.

### Secondary Goal

We also examined program effectiveness as a function of gender, grade, and behavior at the beginning of the school year. The present study included measures of physical and verbal bullying, social exclusion, and malicious gossip in order to avoid specious gender differences (Schafer et al., 2002). Using a similar observation system, Craig and Pepler (1995) found that girls bully just as frequently as boys, although boys report higher levels of bullying than do girls (Olweus, 1993; Whitney et al., 1994). Observational research has also found no age differences in bullying rates within elementary school (Craig et al., 2000).

Evidence of differential program effectiveness is inconclusive. There is some suggestion of greater effectiveness in elementary school than in secondary school (Smith & Ananiadou, 2003) but no reports of age differences within Grades 3 through 6. Reports of gender differences in effectiveness have been inconsistent (Rigby, 2002). For the most part, investigators have indicated no differences (Olweus, 1993) or have not reported any (Stevens et al., 2000; Whitney et al., 1994). Two smaller studies have reported gender-differential effectiveness, one favoring girls (Rahey & Craig, 2002) and the other favoring boys (Eslea & Smith, 1998). We therefore viewed analyses of gender- and age-differential effectiveness as exploratory.

The importance of examining both prevention and intervention effects has been noted in previous literature. Accordingly, we examined possible interactions between group and pretest levels of bullying behavior. Previous observations have shown an overall increase in playground aggression across the school year (Grossman et al., 1997; Reid, Eddy, Fetrow, & Stoolmiller, 1999). We therefore posed two related questions: First, is the program effective at slowing growth in bullying among those who did not engage in bullying during pretest (a preventive effect)? Second, is the program effective at reducing the level of bullying among those observed to bully during the pretest?

## Method

Data were collected as part of a longitudinal study using a cohort-sequential design with a control group at pre- and posttest (Schaie & Baltes, 1977). The current study examined program effects after 1 year of implementation.

### Participants

Six elementary schools in the Pacific Northwest participated in the study. Schools in two suburban districts were matched for size, ethnic

breakdown, and percentage of students receiving free and reduced-price lunches (range = 21%–60%). During the first year of the study (2000–2001), schools in a matched pair in one district were randomly assigned to conditions. In the following year, two matched pairs of schools in another district were randomly assigned to conditions. Criteria for inclusion were that (a) 80% of all staff had to have voted to participate, (b) staff had to have agreed to random assignment to intervention or wait-list control conditions, and (c) principals had to have agreed to refrain from introducing similar interventions during the study.

*Students.* Although all students in Grades 3–6 received the *Steps to Respect* program, their involvement in measurement activities required parental consent. Sixty-four percent ( $n = 1,126$ ) of parents contacted gave consent for their children's participation. We were unable to ascertain if the 36% of children for whom we did not receive parental consent differed from those who participated in the study. Child assent was obtained from fourth to sixth graders during survey administration in the fall. The sample was equally divided by gender (49.4% female). Sample sizes for Grades 3–6 were, respectively, 278, 312, 277, and 259 students. Students in Grades 3 and 4 received Level 1 of the *Steps to Respect* program. Those in Grade 5 and Grade 6 received Levels 2 and 3, respectively.

A subset of 620 students was randomly selected at pretest for playground observation (50.7% male and 49.3% female). This represented 10 students in each control classroom and in each Grade 5 and Grade 6 intervention classroom. In addition, 12 children in each Grade 3 and Grade 4 intervention classroom were selected (comprising a longitudinal sample).

Student ethnic background and English proficiency were reported by teachers. The student sample was 9% African American, 12.7% Asian American, 7.0% Hispanic American, 1.3% Native American, and 70.0% European American. The proportion of students speaking English as a second language (11.5%) did not vary by condition,  $\chi^2(1, N = 1,126) = 0.52, ns$ , nor were there differences in ethnic makeup between conditions,  $\chi^2(4, N = 1,126) = 5.58, ns$ . The ethnic backgrounds and English proficiency of students in the observational sample mirrored those of the larger sample.

*Teachers.* Teachers in 36 experimental and 36 control classrooms completed consent forms. The majority of teachers (84.9%) were female. All teachers agreed to complete study measures, for which they received monetary compensation. Experimental teachers also agreed to bimonthly classroom observations.

### The Program

The *Steps to Respect* program is designed to decrease school bullying problems by (a) increasing staff awareness and responsiveness, (b) fostering socially responsible beliefs, and (c) teaching social-emotional skills to counter bullying and promote healthy relationships. Thus the program also aims to promote skills (e.g., group joining, conflict resolution) associated with general social competence. The *Steps to Respect* program comprises a school-wide program guide, staff training, and classroom lessons for students in Grades 3–6. The program guide presents an overview of curricular content, goals, and research foundations as well as a blueprint for developing school-wide policy and procedures. See Hirschstein and Frey (in press), for a detailed program description.

*Staff training.* The *Steps to Respect* training manual provides a core instructional session for all school staff and two in-depth training sessions for counselors, administrators, and teachers. All staff receive an overview of program goals and key features of program content (e.g., a definition of bullying, a model for responding to bullying reports). Teachers, counselors, and administrators receive additional training in how to coach students involved in bullying. Third- through sixth-grade teachers complete an orientation to classroom materials and instructional strategies.

*Classroom curriculum.* The student curriculum comprises skill and literature-based lessons presented by third- through sixth-grade teachers over a 12–14-week period. Level 1 is taught at Grade 3 or Grade 4, Level

2 at Grade 4 or Grade 5, and Level 3 at Grade 5 or Grade 6. Ten semi-scripted skill lessons focus on social-emotional skills for positive peer relations, emotion management, and recognizing, refusing, and reporting of bullying behavior. Topics include joining groups, distinguishing reporting from tattling, and being a responsible bystander. Instructional strategies include direct instruction, large- and small-group discussions, skills practice, and games. Weekly lessons, totaling about 1 hr, were taught over 2–3 days. Upon completion of skill lessons, teachers implemented a grade-appropriate literature unit, based on existing children's books, which provided further opportunities to explore bullying-related themes.

**Parent engagement.** The program includes a scripted informational overview for parents. Administrators inform parents about the program and the school's antibullying policy and procedures. Finally, take-home letters for parents, provided throughout the classroom curriculum, outline key concepts and skills and describe activities to support their use at home.

### Implementation

**Implementation sequence.** The program was implemented in several phases. First, school bullying prevention teams collaborated with program consultants in the fall to develop the infrastructure to sustain prevention efforts (e.g., handling of reports and coaching for students involved in bullying). Second, school personnel were trained in November. Finally, classroom lessons were implemented in Grades 3–6 from December or January through May.

**Implementation fidelity.** Teachers' ratings of school-wide implementation on a 4-item scale (1 = *poor*, 4 = *excellent*) indicated that by the end of the school year, program policies and procedures were well institutionalized ( $M = 3.25$ ,  $SD = 0.44$ ,  $n = 29$ ). Third- through sixth-grade teachers reported teaching 99.2% of classroom skill lessons. Program consultants recorded bimonthly ratings of observed classroom lesson quality and completion of learning objectives. Kappas based on 50 observational sessions were .62 for lesson quality and .81 for completion of objectives. On a scale of 1 (*poor*) to 3 (*good*), mean lesson quality was rated at 2.24. Observed completion of classroom learning objectives was rated at 91%.

### Study Procedure

**Student surveys and teacher ratings.** Student survey data were collected over a 2-week period in the fall (mid-November) and spring (late April to early May). Surveys were group administered in classrooms, with research personnel reading items aloud and being available to answer student questions one-on-one. Teacher ratings of student interpersonal skills, collected in mid-October and again in May or early June, were hand delivered and retrieved from teachers.

**Playground observations.** Playground observations were collected across 2.5 months in both pre- and posttest periods (October through

December, and April through June). Each child was observed for 5-min sessions approximately once a week over the two 10-week periods, which yielded a subsample of 571 students with pre- and posttest data. Mean observation times were 50 min at each time point. Only children meeting the a priori minimum of 40 min of observation at pre- and posttest ( $n = 544$ ) were included in analyses. Common reasons for incomplete data included a child moving to another school or missing recess.

Observational data were collected using handheld computers and a custom-designed program that runs on a basic PDA (Personal Digital Assistant), recording frequencies, durations, and code-contingent screen changes. Codes were entered by touching a code icon on the screen. Data files were subsequently transferred to a computer.

### Measures

**Teacher ratings of peer interaction skills.** Teachers rated students with the Peer-Preferred Social Behavior subscale of the Walker-McConnell Scale of Social Competence and School Adjustment, Elementary Version (Walker & McConnell, 1995). Reported to have high internal consistency ( $\alpha = .95$ ) and validity, the subscale consists of 17 items measuring social behaviors valued by peers. Using a 5-point scale (1 = *never*, 5 = *frequently*), teachers rated children on items such as "voluntarily provides assistance to peers who require it."

**Student survey of beliefs and behavior.** The Student Experience Survey: What School Is Like for Me is a 60-item measure designed to assess student-reported experiences and attitudes related to bullying. It includes self-reports of behavior, attitudes about behaviors, and perceptions of adults. Some items were adapted from a previous survey developed by WestEd in conjunction with the Committee for Children (Dietsch, Diaz, Frey, & Constantine, 2000). A scale was added to assess student beliefs about the acceptability of bullying and aggression (Erdley & Asher, 1998; Slaby & Guerra, 1988).

An exploratory factor analysis performed on the student survey items yielded eight scales (see Edstrom, Bruschi, & MacKenzie, 2004): Victimization, Direct Bullying/Aggression, Indirect Bullying/Aggression, Acceptance of Bullying/Aggression, Difficulty of Responding Assertively, Bystander Responsibility, Spectator Interest, and Perceived Adult Responsiveness. Seven of the eight scales demonstrated adequate to high internal consistency. (See Table 1 for scale reliability statistics and examples of items.)

**Observational coding.** Drawing on methodology from previous studies of playground aggression (e.g., Grossman et al., 1997; Pepler & Craig, 1995), we designed an observational coding system focused on bullying problems and underlying processes thought to contribute to them (e.g., bystander behaviors, adult intervention, and interpersonal skills). The coding strategy consisted of collecting multiple, continuous focal-individual samples. Each behavior was coded according to a mutually exclusive

Table 1  
Student Experience Survey Scales

Scale	Number of items	$\alpha$	Example item
Self-reported beliefs			
Acceptance of Bullying/Aggression	7	.86	It's okay to say something mean to a kid who really makes you angry.
Bystander Responsibility	5	.88	If my friends were telling lies about another kid I would tell them to stop.
Perceived Adult Responsiveness	4	.59	Adults at my school stop kids from being bullied.
Difficulty of Responding Assertively	5	.81	Kids at school are teasing you. How hard would it be to calmly tell them to stop?
Spectator Interest	6	.88	It's fun to watch other kids being ganged up on.
Self-reported behavior			
Direct Bullying/Aggression	8	.87	I called kids names at school.
Indirect Bullying/Aggression	6	.76	I told my friends to ignore kids that I was mad at.
Victimization	8	.84	A group of kids at school called me mean names.

system of categories (Snell & Frey, 2000). In order to obtain more reliable frequency estimates and normal distributions for low-rate behaviors, several subcategories were combined for analysis in this study. Data were aggregated across 10 sessions to obtain stable means. The five composite codes for focal child behaviors were as follows:

1. *Bullying* included physical, verbal, and indirect aggression involving either (a) a discernible power imbalance between aggressor(s) and target (e.g., a group of children aggressing against a single child) and/or (b) repeated aggression, during the same observation session, by a child toward a nonretaliating peer.<sup>1</sup>
2. *Encouragement of bullying* was coded when focal children laughed or cheered ongoing bullying events. It was also coded for passively watching, in a sustained manner, while an aggressive act took place within 15 feet of the focal child.
3. *Nonbullying aggression* was coded for physical, verbal, or indirect aggression that did not involve a discernible power imbalance or repeated nonreciprocal aggression. Bossy or argumentative behavior was not coded as aggression.
4. *Agreeable social behavior* was coded when a focal child directed neutral or positive acts or statements toward another (e.g., starting a conversation).
5. *Argumentative social behavior* was coded for nonaggressive, negative acts or statements directed toward another child. This category included behaviors such as acting bossy, arguing, or deliberately ignoring another's attempts to enter a group.

In addition, *adult intervention* was coded when any adult intervened in children's behavior during an aggressive bout. Coding of nonaggressive bystander intervention was also attempted. Unfortunately, these variables could not be analyzed because of their very low rates.

*Observer training and agreement.* Two master coders conducted a 200-hr training of 13 coders that included code memorization, drills in PDA mechanics, and review and practice coding of videotapes of children on playgrounds. Coders were required to meet a criterion of  $\kappa = .70$  on increasingly difficult tape segments, and then on field events, prior to collecting data. To reduce child reactivity, observers were minimally responsive and started their observations during a habituation period 2 weeks prior to actual data collection.

Random agreement checks were made by master coders for 15% of the sessions on an event-by-event basis in single observation sessions (overall Cohen's  $\kappa = .80$ ). In order to provide a more stringent test of reliability, separate agreement statistics were also calculated (bullying = .63, nonbullying aggression = .54, encouragement of bullying = .55, agreeable social behavior = .81, and argumentative social behavior = .56). Kappas for individual codes were in the fair range (Fleiss, 1981) and consistent with those obtained in previous observational research on aggressive behavior (e.g., Dodge, Coie, Pettit, & Price, 1990; Grossman et al., 1997). Actual reliabilities of the codes may be higher than these numbers suggest. Whereas these kappas were computed on an event-by-event basis within single sessions, the analyzed behavior levels were aggregated over a minimum of 10 sessions (Hubbard, Dodge, Cillessen, Coie, & Schwartz, 2001). A final analysis to examine the ability of coders to distinguish between bullying and nonbullying aggression indicated excellent discriminability ( $\kappa = .80$ ).

## Results

In this section, we first present data pertaining to retention rates and group differences in pretest levels as a function of retention status. Second, we provide the results of hierarchical linear anal-

yses used to test (a) teacher and student-reported skills, beliefs, and behaviors and (b) changes in observed playground behavior. Finally, we briefly note overall grade and sex differences in skills, beliefs, and behaviors.

### Student Retention

There were no group differences in retention rates (91.3%) from pre- to posttest,  $\chi^2(1, N = 1,126) = 0.04, ns$ . Of the 1,023 students retained at posttest, 4.9% ( $n = 51$ ) were unavailable for surveying at either the pretest or posttest. Owing to participants' failure to respond to specific survey items, sample sizes were further reduced (range = 907 to 919) on individual scales. Comparisons of pretest student surveys and teacher ratings revealed neither any significant group or retention status differences (all  $F_s < 1$ ) nor any group by retention status interactions (all  $F_s < 1.09$ ).

Retention rates found for the observation subsample (92.4%) also did not differ by group,  $\chi^2(1, N = 620) = 1.72, ns$ . Pretest behaviors of students in the final subsample were compared with those of all other students for whom we had at least 40 min of observation time at pretest. There was one significant interaction of group and retention status,  $F(1, 619) = 5.78, p < .05$ , for encouragement of bullying. Follow-up tests revealed a significant group difference at pretest in the original sample ( $p < .05$ ) but not in the final sample of 544 students ( $F < 1$ ).

### Analytic Strategy

Intraclass correlations performed with classroom as a random factor were significant or nearly significant for 6 of 13 dependent measures. For consistency, we performed hierarchical linear analyses for all tests of group effects, using SPSS Linear Mixed Models 12.0. Two-level models examined student outcomes in terms of individual- and classroom-level variables while concurrently adjusting for shared error due to classroom nesting (Hedeker, Gibbons, & Flay, 1994).

### Reported Beliefs, Behaviors, and Skills

Scale-item means and standard deviations for student self-reports and teacher ratings of peer interaction skills are shown in Table 2. Because of low variability in scores, the Spectator Interest scale was dropped from further consideration.

To improve normality of distributions, we performed square root transformations on all self-report measures save the perceived Difficulty of Responding Assertively scale. The basic model included the intercept and tested for the fixed effects of group and three covariates: gender, grade level, and fall pretest scores. With the exception of Difficulty of Responding Assertively, no significant or near-significant ( $p < .10$ ) interactions of group with covariates were found.

*Beliefs related to bullying.* As shown in Table 2, predicted overall group effects were found for three of four bullying-related

<sup>1</sup> Previous research indicates that most aggression occurring between asymmetrically aggressive dyads is proactive, with one child repeatedly targeted for aggression (Dodge, Price, Coie, & Christopoulos, 1990). Although bullying is not limited to proactive aggression, a substantial percentage of proactive aggression is bullying (Coie & Dodge, 1998).

Table 2  
*Reported Beliefs, Behaviors, and Skills: Pre- and Posttest Item Means and Tests of Group Effects*

Variable	Fall pretest				Spring posttest				Group effect
	Intervention		Control		Intervention		Control		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Self-reported beliefs									
Acceptance of bullying/aggression <sup>a</sup>	0.78	0.76	0.80	0.79	0.76	0.79	0.87	0.81	$F(1, 73.8) = 8.51^{**}$
Bystander responsibility <sup>b</sup>	2.39	0.72	2.34	0.70	2.32	0.75	2.18	0.80	$F(1, 93.3) = 3.93^*$
Perceived adult responsiveness <sup>b</sup>	2.13	0.68	2.13	0.63	2.12	0.70	2.03	0.67	$F(1, 93.2) = 5.30^*$
Difficulty of responding assertively <sup>c</sup>	1.08	0.78	1.09	0.77	1.04	0.81	1.13	0.81	$F(1, 63.0) = 3.14^\dagger$
Self-reported behaviors									
Direct aggression <sup>d</sup>	0.46	0.59	0.56	0.66	0.48	0.62	0.62	0.71	$F(1, 68.7) = 2.05$
Indirect aggression <sup>d</sup>	0.88	0.72	0.94	0.73	0.90	0.74	0.96	0.83	$F < 1$
Victimization <sup>d</sup>	1.01	0.79	1.07	0.82	0.90	0.82	1.01	0.83	$F(1, 72.4) = 3.74^\dagger$
Teacher-rated interaction skills	3.72	0.90	3.71	0.87	3.82	0.87	3.81	0.86	$F < 1$

<sup>a</sup> Range = 0 (*don't agree*) to 3 (*agree a lot*). <sup>b</sup> Range = 0 (*not true*) to 3 (*very true*). <sup>c</sup> Range = 0 (*not hard at all*) to 3 (*really hard*). <sup>d</sup> Range = 0 (*never*) to 4 (*more than once a week*).  
 $^\dagger p < .10$ .  $^* p < .05$ .  $^{**} p < .01$ .

beliefs, and the fourth approached significance. At posttest, students in intervention schools were less accepting of bullying/aggression, felt more responsibility to intervene with friends who bullied, and reported greater adult responsiveness than did those in control schools. Means indicated the differences were due to a deterioration of beliefs among control-group students.

An overall group difference in the perceived difficulty of responding assertively to bullying approached significance, modified by a Group  $\times$  Grade interaction,  $F(1, 63.8) = 5.51, p < .05$ , and a near-significant Group  $\times$  Gender interaction,  $F(1, 913.6) = 3.67, p < .10$ . Estimated means and confidence intervals are presented in Table 3. As illustrated by the nonoverlapping intervals, most intervention students in Grades 5 and 6 perceived the difficulty to be lower than did their peers in control schools ( $p < .05$ ). Similarly, boys in the intervention group tended to regard assertive responses as less difficult than did boys in the control group ( $p < .06$ ). The overlapping intervals for girls showed no differences in perceived difficulty attributable to the intervention.

*Reported behavior and teacher-rated skills.* As shown in Table 2, students in the intervention group tended to report less victimization at posttest than did those in the control group. No group differences were found for student self-reported bullying/aggression. Contrary to predictions, teacher ratings of interaction skill also did not vary by group.

#### Playground Observations

*Data screening.* Chi-square analyses indicated that participation in bullying behavior was reasonably well distributed and did not vary by group. Of the 544 students, 60.7% bullied, 47.8% encouraged bullying, 75.4% engaged in nonbullying aggression, and 56.4% were targeted for bullying during either the pre- or posttest observations.

Means and standard deviations for behaviors at pretest and posttest are presented in Table 4. Unlike participation rates, frequency rates of antisocial behaviors were severely nonnor-

Table 3  
*Perceived Difficulty of Responding Assertively to Bullying: Group Interactions With Grade Level and Gender*

Grade level or gender and group	<i>n</i>	<i>M</i> <sup>a</sup>	<i>SE</i>	95% confidence interval	
				Lower bound	Upper bound
Grades 3 and 4					
Intervention	305	1.11	0.04	1.02	1.20
Control	285	1.08	0.05	0.99	1.18
Grades 5 and 6					
Intervention	244	0.95	0.05	0.84	1.06
Control	292	1.15	0.05	1.06	1.24
Boys					
Intervention	272	0.97	0.05	0.88	1.07
Control	298	1.15	0.04	1.06	1.24
Girls					
Intervention	277	1.09	0.05	1	1.19
Control	279	1.09	0.05	0.99	1.18

<sup>a</sup> Estimated means adjusted for pretest scores.

mal at pre- and posttest. Examining several possible data transformations, we found that change scores, calculated as the difference between fall pretest and spring posttest frequencies, provided acceptable distributions and residual values after outliers in the dependent measures (8 control cases, 9 intervention cases) were brought in to a value that was 0.2 standard deviations more extreme than the next attached value or 3.29 standard deviations, whichever was more extreme (Tabachnick & Fidell, 1996). Statistically comparable to performing a repeated measures analysis, change scores provide an unbiased estimate of true change regardless of the baseline value (Zumbo, 1999). Previous investigations of playground aggression have used change scores to address distribution problems typical of low-frequency behaviors (e.g., Grossman et al., 1997; Reid et al., 1999).

We used the occurrence or nonoccurrence of the specific problem behavior at pretest as a factor to test specifically whether the program was effective at (a) reducing problem behaviors among students who had exhibited them at pretest and (b) maintaining low levels among those who had not. Use of a dichotomous variable also enabled us to circumvent problems with nonnormality in the pretest variables. Because most children exhibited some agreeable or argumentative behavior, we split frequencies at the median for these two variables in order to provide a design analogous to that for bullying and aggression.

Hierarchical modeling of observed behavior change included the intercept. We tested a factorial model (group and pretest occurrence of the dependent variable) with gender and grade level entered as covariates. Interactions of group with the covariates were included in the model whenever preliminary analyses showed significant or near-significant ( $p < .10$ ) effects.

*Behaviors related to bullying and aggression.* A significant main effect for change in bullying was qualified by a near-significant interaction of group and baseline occurrence,  $F(1, 541) = 3.20, p < .10$ . Table 5 indicates that the overall effect was primarily due to reductions among those who bullied at pretest ( $p < .01$ ), with virtually no overlap between confidence intervals of the intervention and control groups. Substantial overlap was found among those who were not observed to bully at pretest, indicating the need to discount the small mean difference.

There was a near-significant decline in bystander encouragement of bullying among intervention students relative to control students. Although the interaction of group with pretest occurrence  $F(1, 541) = 2.18$ , was not significant, Table 5 shows a pattern of results similar to that for bullying: virtually no overlap in confidence intervals for intervention- and control-group students who encouraged bullying during pretest observations. Low rates of bystander encouragement at pretest (see Table 5) may have reduced our opportunity to find a significant overall group difference.

Contrary to predictions, victimization by bullying showed no significant group differences. Nonbullying aggression showed extremely high variability despite transformation via change scores. Group differences in nonbullying aggression were not significant.

*Changes in argumentative and agreeable behavior.* As shown in Table 4, children in the intervention group displayed decreased argumentative behavior, with virtually no decrease in the control group. Changes in agreeable behavior also varied significantly by group. Table 6 displays estimated means and confidence intervals relative to a near-significant Group  $\times$  Gender interaction,  $F(1, 477) = 3.49, p < .10$ . Boys in the intervention group tended to show increases in agreeable interactions, with confidence intervals that did not overlap with those for boys in the control group. Intervention-group girls showed little program benefit with respect to agreeable behavior, overlapping almost completely with control-group girls.

#### Overall Grade and Gender Differences

Observed bullying, bystander encouragement, and aggression did not vary overall by grade level or gender. Younger students, however, were targeted for bullying more frequently ( $p < .01$ ) and reported more victimization than did older students ( $p < .05$ ). Older students reported more direct and indirect aggression, more acceptance of bullying/aggression, and less responsibility to intervene with friends who bully than did younger students (all  $ps < .01$ ). Older students did credit adults with greater responsiveness regarding bullying than did their younger counterparts ( $p < .01$ ).

Girls were rated as more socially skilled than boys by their teachers ( $p < .05$ ). Compared with boys, girls reported more indirect aggression but also less acceptance of bullying/aggression and more responsibility to intervene with friends who bully (all

Table 4  
*Observed Behavior: Pretest and Posttest Means and Tests of Group Effects*

Observed behavior	Fall pretest				Spring posttest				Group effect
	Intervention		Control		Intervention		Control		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Aggressive behavior (mean no. per hour)									
Bullying	0.85	1.47	0.73	1.21	0.97	1.71	1.19	2.11	$F(1, 91.3) = 5.02^*$
Encourage bullying	0.58	1.05	0.65	1.31	0.49	1.01	0.78	1.62	$F(1, 75.8) = 3.24^\dagger$
Target of bullying	0.58	1.14	0.59	0.95	0.80	1.51	0.86	1.44	$F < 1$
Nonbullying aggression	1.86	2.60	1.87	2.97	1.60	2.66	1.86	2.75	$F(1, 89.2) = 1.00, ns$
Social interaction (percentage of time)									
Agreeable social	41.48	15.86	41.81	15.77	42.78	16.79	41.26	16.57	$F(1, 261.9) = 4.99^*$
Argumentative social	3.25	3.39	2.98	2.59	2.60	2.61	2.80	2.81	$F(1, 216.1) = 4.84^*$

Note. Tests of group effects were based on the difference between pre- and posttest levels.

$^\dagger p < .10$ .  $^* p < .05$ .

Table 5  
*Estimated Mean Changes in Bullying Behavior as a Function of Group and Pretest Occurrence*

Pretest behavior and group	<i>n</i>	<i>M</i> <sup>a</sup>	<i>SE</i>	95% confidence interval	
				Lower bound	Upper bound
Pretest bullying					
Intervention	110	-0.99	0.19	-1.37	-0.61
Control	99	-0.26	0.20	-0.66	0.14
No pretest bullying					
Intervention	186	0.73	0.16	0.42	1.05
Control	149	0.92	0.17	0.59	1.26
Pretest encouragement of bullying					
Intervention	89	-1.38	0.16	-1.70	-1.06
Control	75	-0.88	0.18	-1.23	-0.53
No pretest encouragement of bullying					
Intervention	207	0.43	0.13	0.17	0.69
Control	173	0.58	0.14	0.31	0.85

<sup>a</sup> Estimated means adjusted for gender and grade level.

$p < .01$ ). Boys believed adults to be more responsive to bullying than did girls ( $p < .01$ ).

### Discussion

Using a rigorous experimental design and analytic strategy, this evaluation provided evidence of positive program effects with respect to (a) observed bullying behavior, (b) observed social interaction, and (c) attitudes related to bullying. These effects were largely consistent across gender and grade.

#### *Playground Behavior*

*Bullying and destructive bystander behavior.* Effect sizes calculated for these six schools showed modest effects overall, effects that are likely to differ with implementation quality (Edstrom, Hirschstein, Frey, Snell, & MacKenzie, 2004). Student characteristics were also related to the magnitude of some program effects. Relative decreases in bullying behavior were found largely among those who had engaged in bullying during the pretest, indicating an intervention effect. Standardized mean differences (Cohen, 1988)

calculated for the six schools in this study illustrate the relative magnitudes of the intervention and prevention effects ( $d = .31$  and  $d = .15$ , respectively).

A similar, but less robust, pattern ( $d = .24$  and  $d = .14$ , respectively) was found for bystander encouragement of bullying, perhaps because of a lower base rate at pretest. It is important to note that this variable represents bystander behaviors observed in focal children only, because observers could not simultaneously code focal bullying participants and nonfocal bystanders. Therefore, a low base rate does not indicate that bystanders were seldom in attendance during bullying events.

In fact, these observations provide objective evidence of the widespread nature of elementary school students' participation in bullying behavior. Although not all students bullied or encouraged others to bully, the majority (77.0%) were observed doing so at one or both time points. This prevalence has an important practical significance. When large numbers of young people participate in or are exposed to bullying, even a modestly effective intervention can have large consequences for students within a school (Abelson, 1985).

To illustrate, we extrapolated from the mean levels of observed bullying. If children are on the playground for 60 min a day, then a control classroom of 25 children would perpetrate 30 acts of playground bullying on an average spring day. In contrast, 24 acts of playground bullying would be perpetrated by children in an intervention classroom. Similarly, bystanders in a control classroom would watch and encourage another 20 bullying events per day.<sup>2</sup> In contrast, students in each intervention classroom would encourage bullying 13 times per day. Together, these two measures represent 24.6% fewer bullying behaviors among intervention students. If we further extrapolate to 72 days in the posttest period (March to mid-June), we would find about 900 fewer

Table 6  
*Estimated Mean Changes in Agreeable Behavior by Gender and Group*

Group	<i>n</i>	<i>M</i> <sup>a</sup>	<i>SE</i>	95% confidence interval	
				Lower bound	Upper bound
Boys					
Intervention	150	2.10	1.22	-0.29	4.49
Control	129	-0.91	1.31	-3.49	1.67
Girls					
Intervention	146	0.38	1.23	-2.04	2.80
Control	119	0.03	1.36	-2.65	2.71

<sup>a</sup> Estimated means adjusted for grade level.

<sup>2</sup> Typically, there were no more than two observers on the playground at any one time, making it extremely unlikely that one would be observing focal bullying while another observed a focal bystander to the same event.

bullying behaviors perpetrated by students in intervention classrooms than in control classrooms.

Positive results in observed behavior were not mirrored by reductions in self-reported bullying/aggression. Although self-reports may reflect events not accessible to observers (Pellegrini & Bartini, 2000), it is also possible that the intervention may have sensitized students to the bullying they perpetrated and experienced (Smith & Ananiadou, 2003), contributing to the null findings.

*Observed victimization.* Support for this possibility comes from a study of classroom implementation factors, which provides a context for interpreting the near-significant reduction in victimization among the intervention group relative to the control group. Within the intervention group, self-reported victimization was actually higher at posttest when the quality of the *Steps to Respect* lessons was rated more highly by observers (Edstrom, Hirschstein, et al., 2004). Arguably, clear and engaging lessons would be most likely to facilitate student learning and may have succeeded in expanding students' definitions of bullying and victimization. Further assistance in interpreting this finding awaits evaluation after a longer implementation period.

Very low baseline rates of observed victimization may have contributed to the lack of group differences on that measure. A puzzling finding was that children were observed to bully more frequently than they were observed being victimized. A possible explanation is that the presence of a nearby observer had differential effects. Nonfocal children interested in bullying others could easily choose a target out of proximity to the observer, resulting in less bullying of focal children. A focal child would have had to exercise more self-restraint in order to avoid having his or her aggression recorded throughout an entire observation session, however. The use of audio- or video-recorded interactions to observe behavior (e.g., Craig & Pepler, 1995; Tapper & Boulton, 2002) might have helped determine whether the discrepancy was indeed due to observer reactivity.

### *Bullying-Related Beliefs*

In the current study, we found that control-group students became more accepting of bullying and aggression than did intervention-group students across the school year, a deterioration that was not accompanied by improved attitudes in the intervention group. There appeared to be an erosion of control students' sense of responsibility to stop friends from bullying and of their perceptions of adult responsiveness. Such declines might encourage children in a control school to expect fewer negative consequences (e.g., adult sanctions or peer disapproval) for bullying and to feel fewer inhibitions against it as the school year progresses, in accordance with a transactional model of normative beliefs and aggressive behavior (Egan, Monson, & Perry, 1998; Huesmann & Guerra, 1997). We note, however, that the magnitude of program effects on bullying beliefs was quite small (*ds* ranging from .16 to .19). Although it is possible that small changes in the beliefs of many children may contribute to further deteriorations in behavior, it is also possible that during such a short time span, attitudes and behaviors decline in parallel, without a causal relationship.

### *Social-Emotional Skills*

Teacher ratings did not indicate an increase in peer interaction skills. This may reflect the program emphasis. Intervention materials focus relatively more attention on skills to counter bullying than on general friendship and conflict resolution skills. An alternate explanation concerns methodology: Teachers have limited venues for observing increased peer interaction skills, compared with playground observers (Pellegrini & Bartini, 2000), who are privy to behaviors occurring in an unstructured, peer-dominated setting. In this setting, and in proximity to nonreactive observers, students were perhaps less likely to inhibit behaviors that teachers might censure. Playground observations in this study showed a small relative decline in argumentative, bossy behavior ( $d = .13$ ) and more agreeable interactions among intervention-group boys ( $d = .22$ ), changes suggesting incremental increases in social competence that might well have escaped teacher notice.

### *Gender and Grade Effects*

Compared with girls, boys benefited more from program participation in two respects: They showed increases in agreeable behavior and a somewhat greater decline in the perceived difficulty of responding assertively to bullying ( $d = .25$ ) relative to boys in the control group. Girls did not differ from their counterparts in the control group on these measures. Overall, girls were less accepting of bullying and aggression but indicated they were more indirectly aggressive than boys.

The sole example of age-differential effectiveness favored older students, whose program participation reduced the perceived difficulty of responding to bullying in a calm, controlled manner ( $d = .28$ ). Responding effectively to bullying requires a strong, confident delivery, which may be easier to muster among older students. Overall, lower rates of victimization in the higher grades may be indicative of greater confidence, size, or skill. Self-reports of aggression and attitudes supporting aggression were higher in the older students, confirming age trends noted in previous research (Pellegrini & Long, 2002; Rigby & Slee, 1991).

### *Caveats and Implications for Research*

It should be noted that grade was confounded in this study with program level. Moreover, older students participating in the study had not received previous levels, as they would have under recommended implementation conditions. Longitudinal work is needed to examine the cumulative impacts of bullying prevention for more than a 4-month period.

Our choice to conduct microanalytic observations in a small number of schools limited our ability to analyze changes in school-wide procedures and systems, factors thought to be particularly important in bullying prevention (Olweus, 1993, 1999; Pepler et al., 1999; Smith & Sharp, 1994). An important question, for example, is the extent to which schools provided intensive, individually focused interventions with children needing more direct guidance than that provided by the universal aspects of the program. Also, survey measures of playground monitoring (e.g., Leff, Power, Costigan, & Manz, 2003) might shed light on what appears to be an extremely low level of adult intervention in bullying.

A third caveat concerns the use of self-report measures that describe specific behaviors rather than use the term *bullying*. Use of the word *bully* may elicit socially desirable responses (Espelage & Swearer, 2003). Nevertheless, previous survey research using the term (e.g., Olweus, 1999; Whitney et al., 1994) has enabled investigators to amass a significant body of work that shares a common methodology. Further work is needed to examine the strengths and limitations of each approach.

A final weakness, and strength, is the *in vivo* coding of bullying. This study demonstrates that playground coders can reliably distinguish between bullying and nonbullying aggression. It is nevertheless likely that we failed to observe all bullying behaviors. Gossip, for example, can be difficult to hear on the playground. Observations using recording devices may be especially well-suited for capturing indirect aggression (e.g., Pepler & Craig, 1995; Tappet & Boulton, 2002).

### Conclusion

A unique contribution of the current random control trial is the use of unbiased observations to measure verbal, physical, and relational bullying behavior on the playground. We know of no other bullying prevention evaluation that includes field observations as part of a multimethod strategy. Given indications that direct observations may be equally or more sensitive to actual behavioral change than adult reports (Patterson, 1982) or peer reports (Hymel, 1986), evaluators may be encouraged to undertake observations more frequently. An additional benefit is the possibility of increasing our understanding of bullying dynamics, as demonstrated by the substantial contributions of previous observational research (e.g., Craig et al., 2000; Hanish, Ryan, Martin, & Fabes, 2005; Pellegrini & Bartini, 2000).

Implementation of the *Steps to Respect* program resulted in positive changes in observed playground bullying, normative beliefs, and social interaction skills. Notably, both bullying and the attitudes believed to support its execution were reduced, relative to the control group, within a relatively short period of time. The trend toward reductions in bystander encouragement of playground bullying was also heartening, as reducing the number of children who provide an audience and incitement to bullying may yield additional benefits in subsequent years (Salmivalli, 1999). Bystander skills, beliefs, and behavior represent frequently overlooked aspects of violence prevention (Espelage, Holt, & Henkel, 2003; Farmer et al., 2002) and were a particular focus of the program. If schools are able to alter peer norms and behavior, increase student skills, and sustain adult prevention efforts, the positive effects of their work may gather momentum and strengthen over time.

### References

- Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129–133.
- Andershed, H., Kerr, M., & Stattin, H. (2001). Bullying in school and violence on the streets: Are the same people involved? *Journal of Scandinavian Studies in Criminology and Crime Prevention*, *2*, 31–49.
- Boulton, M. J., Trueman, M., Chau, C., Whitehand, C., & Amatya, K. (1999). Concurrent and longitudinal links between friendship and peer victimization: Implications for befriending interventions. *Journal of Adolescence*, *22*, 461–466.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coie, J. D., & Dodge, K. A. (1998). The development of aggression and antisocial behavior. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 779–862). New York: Wiley.
- Committee for Children. (2001). *Steps to Respect: A bullying prevention program*. Seattle, WA: Author.
- Connolly, J., Pepler, D., Craig, W., & Taradash, A. (2000). Dating experiences of bullies in early adolescence. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, *5*, 299–310.
- Craig, W. M., & Pepler, D. J. (1995). Peer processes in bullying and victimization: An observational study. *Exceptionality Education Canada*, *5*, 81–95.
- Craig, W. M., Pepler, D., & Atlas, R. (2000). Observations of bullying in the playground and in the classroom [Special Issue: Bullies and victims]. *School Psychology International*, *21*, 22–36.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383.
- Dietsch, B. J., Diaz, M., Frey, K. S., & Constantine, N. (2000). *Beliefs and Behaviors Survey*. Los Alamitos, CA: WestEd.
- Dodge, K. A., Coie, J. D., Pettit, G. S., & Price, J. M. (1990). Peer status and aggression in boys' groups: Developmental and contextual analyses. *Child Development*, *61*, 1289–1309.
- Dodge, K. A., Price, J. M., Coie, J. D., & Christopoulos, C. (1990). On the development of aggressive dyadic relationships in boys' peer groups. *Human Development*, *33*, 260–270.
- Edstrom, L. V., Bruschi, C. J., & MacKenzie, E. P. (2004). *The Student Experience Survey* (Tech. Rep. 3.1). Seattle, WA: Committee for Children.
- Edstrom, L. V., Hirschstein, M., Frey, K., Snell, J. L., & MacKenzie, E. P. (2004). *Classroom level influences in school-based bullying prevention: Key program components and implications for instruction*. Paper presented at the annual meeting of the Society for Prevention Research, Quebec City, Quebec, Canada.
- Egan, S. K., Monson, T. C., & Perry, D. G. (1998). Social-cognitive influences on change in aggression over time. *Developmental Psychology*, *34*, 996–1006.
- Eisenberg, N., & Fabes, R. (1998). Prosocial development. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 701–778). New York: Wiley.
- Endresen, I. M., & Olweus, D. (2001). Self-reported empathy in Norwegian adolescents: Sex differences, age trends, and relationship to bullying. In A. C. Bohart, C. Arthur, & D. J. Stipek (Eds.), *Constructive & destructive behavior: Implications for family, school, and society* (pp. 147–165). Washington, DC: American Psychological Association.
- Erdley, C. A., & Asher, S. R. (1998). Linkages between children's beliefs about the legitimacy of aggression and their behavior. *Social Development*, *7*, 321–339.
- Eslea, M., & Smith, P. K. (1998). The long-term effectiveness of anti-bullying work in primary schools. *Educational Research*, *40*, 203–218.
- Espelage, D. L., Bosworth, K., & Simon, T. R. (2001). Short-term stability and change of bullying in middle school students: An examination of demographic, psychosocial, and environmental correlates. *Violence and Victims*, *16*, 411–426.
- Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development*, *74*, 205–220.
- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and

- victimization: What have we learned and where do we go from here? *School Psychology Review*, 32, 365–383.
- Farmer, T. W., Leung, M.-C., Pearl, R., Rodkin, P. C., Cadwallader, T. W., & Van Acker, R. (2002). Deviant or diverse peer groups? The peer affiliations of aggressive elementary students. *Journal of Educational Psychology*, 94, 611–620.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.
- Grossman, D. C., Neckerman, H. J., Koepsell, T. D., Liu, P. Y., Asher, K. N., Beland, K., Frey, K., & Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association*, 277, 1605–1611.
- Hanish, L. D., Ryan, P., Martin, C. L., & Fabes, R. A. (2005). The social context of young children's peer victimization. *Social Development*, 14, 2–19.
- 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 and Allied Disciplines*, 41, 441–455.
- Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development*, 10, 512–527.
- Hedeker, D., Gibbons, R. D., & Flay, B. R. (1994). Random-effects regression modeling for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62, 757–765.
- Hirschstein, M. K., & Frey, K. S. (in press). Promoting behaviors and beliefs that reduce bullying: The *Steps to Respect* program. In S. R. Jimerson & M. J. Furlong (Eds.), *The handbook of school violence and school safety: From research to practice*. Mahwah, NJ: Erlbaum.
- Hoover, J. H., Oliver, R., & Hazler, R. J. (1992). Bullying: Perceptions of adolescent victims in the Midwestern USA. *School Psychology International*, 13, 5–16.
- Hubbard, J. A., Dodge, K. A., Cillessen, A. H. N., Coie, J. D., & Schwartz, D. (2001). The dyadic nature of social information processing in boys' reactive and proactive aggression. *Journal of Personality and Social Psychology*, 80, 268–280.
- Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, 72, 408–419.
- Hymel, S. (1986). Interpretations of peer behavior: Affective bias in childhood and adolescence. *Child Development*, 57, 431–445.
- Kaukiainen, A., Salmivalli, C., Lagerspetz, K., Tamminen, M., Vauras, H. M., & Postkiparta, E. (2002). Learning difficulties, social intelligence, and self concept: Connections to bully-victim problems. *Scandinavian Journal of Psychology*, 43, 269–278.
- Kochenderfer, B. J., & Ladd, G. W. (1997). Victimized children's responses to peers' aggression: Behaviors associated with reduced versus continued victimization. *Development and Psychopathology*, 9, 59–73.
- Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development*, 74, 1344–1367.
- Leff, S. S., Power, T. J., Costigan, T. E., & Manz, P. H. (2003). Assessing the climate of the playground and lunchroom: Implications for bullying prevention programming. *School Psychology Review*, 32, 418–430.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., 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.
- O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence*, 22, 437–452.
- Olweus, D. (1993). *Bullying at school*. Cambridge, MA: Blackwell.
- Olweus, D. (1999). Sweden, Norway. In P. K. Smith, Y. Morita, J. Junger-Tas, D. Olweus, R. F. Catalano, & P. Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp. 7–48). New York: Routledge.
- Owens, L., Slee, P., & Shute, R. (2000). "It hurts a hell of a lot. . .": The effects of indirect aggression on teenage girls. *School Psychology International*, 21, 359–376.
- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Pellegrini, A. D. (2002). Bullying, victimization, and sexual harassment during the transition to middle school. *Educational Psychologist*, 37, 151–163.
- Pellegrini, A. D., & Bartini, M. (2000). An empirical comparison of methods of sampling aggression and victimization in school settings. *Journal of Educational Psychology*, 92, 360–366.
- Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Developmental Psychology*, 20, 259–280.
- Pepler, D. J., & Craig, W. M. (1995). A peek behind the fence: Naturalistic observations of aggressive children with remote audiovisual recording. *Developmental Psychology*, 31, 548–553.
- Pepler, D., Craig, W. M., & O'Connell, P. (1999). Understanding bullying from a dynamic systems perspective. In A. Slater & D. Muir (Eds.), *The Blackwell reader in developmental psychology* (pp. 440–451). Oxford, England: Blackwell.
- Perry, D. G., Hodges, E. V. E., & Egan, S. K. (2001). Determinants of chronic victimization by peers: A review and new model of family influence. In J. Juvonen & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 73–104). New York: Guilford Press.
- Rahey, L., & Craig, W. M. (2002). Evaluation of an ecological program to reduce bullying in schools. *Canadian Journal of Counselling*, 36, 281–296.
- Reid, J. B., Eddy, J. M., Fetrow, R. A., & Stoolmiller, M. (1999). Description and immediate impacts of a preventive intervention for conduct problems. *American Journal of Community Psychology*, 27, 483–517.
- Rigby, K. (2002). *A meta-evaluation of methods and approaches to reducing bullying in pre-schools and in early primary school in Australia*. Canberra, Australia: Commonwealth Attorney-General's Department.
- Rigby, K., & Slee, P. T. (1991). Bullying among Australian school children: Reported behavior and attitudes toward victims. *Journal of Social Psychology*, 131, 615–627.
- Roland, E. (2000). Bullying in school: Three national innovations in Norwegian schools in 15 years [Special Issue: Bullying in the schools]. *Aggressive Behavior*, 26, 135–143.
- Salmivalli, C. (1999). Participant role approach to school bullying: Implications for intervention. *Journal of Adolescence*, 22, 453–459.
- Schafer, M., Werner, N. E., & Crick, N. R. (2002). A comparison of two approaches to the study of negative peer treatment: General victimization and bully/victim problems among German schoolchildren. *British Journal of Developmental Psychology*, 20, 281–306.
- Schaie, K. W., & Baltes, P. B. (1977). On sequential strategies in developmental research: Description or explanation. *Human Development*, 18, 384–390.
- Schwartz, D. (2000). Subtypes of victims and aggressors in children's peer groups. *Journal of Abnormal Child Psychology*, 28, 181–192.
- Schwartz, D., Dodge, K. A., & Coie, J. D. (1993). The emergence of chronic peer victimization in boys' play groups. *Child Development*, 64, 1755–1772.
- Slaby, R. G., & Guerra, N. G. (1988). Cognitive mediators of aggression in adolescent offenders: I. Assessment. *Developmental Psychology*, 24, 580–588.
- Smith, P. K., & Ananiadou, K. (2003). The nature of school bullying and the effectiveness of school-based interventions. *Journal of Applied Psychoanalytic Studies*, 5, 189–209.

- Smith, P. K., & Brain, P. (2000). Bullying in schools: Lessons from two decades of research [Special Issue: Bullying in the schools]. *Aggressive Behavior, 26*, 1–9.
- Smith, P. K., & Sharp, S. (1994). *Tackling bullying in your school: A practical handbook for teachers*. New York: Routledge.
- Snell, J. L., & Frey, K. S. (2000). *Playground aggressors, targets and bystanders (PATAB)*. Unpublished training manual, Committee for Children, Seattle, WA.
- Snell, J., MacKenzie, E., & Frey, K. (2002). Bullying prevention in elementary schools: The importance of adult leadership, peer group support, and student social-emotional skills. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 351–372). Bethesda, MD: National Association of School Psychologists.
- Stevens, V., Van Oost, P., & de Bourdeaudhuij, I. (2000). The effects of an anti-bullying intervention programme on peers' attitudes and behaviour. *Journal of Adolescence, 23*, 21–34.
- Sutton, J., Smith, P. K., & Swettenham, J. (1999). Bullying and “theory of mind”: A critique of the “social skills deficit” view of anti-social behaviour. *Social Development, 8*, 117–127.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins College.
- Tapper, K., & Boulton, M. J. (2002). Studying aggression in school children: The use of a wireless microphone and micro-video camera. *Aggressive Behavior, 28*, 356–365.
- Walker, H. M., & McConnell, S. R. (1995). *Walker–McConnell Scale of Social Competence and School Adjustment: Elementary version*. San Diego, CA: Singular.
- Whitney, I., Rivers, I., Smith, P. K., & Sharp, S. (1994). The Sheffield project: Methodology and findings. In P. K. Smith & S. Sharp (Eds.), *School bullying: Insights and perspectives* (pp. 20–56). London: Routledge.
- Zumbo, B. D. (1999). The simple difference score as an inherently poor measure of change: Some reality, much mythology. *Advances in Social Science Methodology, 5*, 269–304.

Received February 17, 2004

Revision received November 17, 2004

Accepted November 21, 2004 ■