

## ABSTRACT

Title of Document:                   OUTCOMES OF AN ELEMENTARY  
  GRADES SOCIAL COMPETENCE  
  EXPERIMENT ACCORDING TO STUDENT  
  SELF-REPORT

Elise T. Harak, Doctor of Philosophy. 2008

Directed By:                           Professor Gary D. Gottfredson  
  Counseling and Personnel Services

Problem behaviors that emerge in early childhood often persist through adolescence. Evaluations provide evidence that social skills programs in elementary schools can reduce student aggression. There is some evidence that social skills programs also increase social skills, academic commitment, and achievement. Outcome evaluations have more often focused on aggression than on social skills and academics, however. The present study is a randomized, controlled trial evaluating the effects of one popular social skills instructional program, Second Step, in six treatment and six control schools after two years of implementation. Despite the widespread use of Second Step, few evaluations have assessed its effects. The existing evaluations have either: (a) lacked randomization, (b) had small samples, (c) not measured implementation, or (d) were implemented for one year or less. In the present evaluation, implementation data were collected from all teachers as each lesson was

completed. Overall implementation was high across two years. Treatment effects were assessed on nine self-report measures including Engagement in Learning, prosocial behaviors (Altruism, Empathy, and Self-Restraint) and problem behaviors and attitudes (Rebellious Behavior, Aggression, Victimization, Acceptability of Aggression, and Hostile Attribution Bias). Analyses completed using hierarchical linear modeling (HLM) implied that treatment did not statistically significantly affect individual student self-reports net of individual characteristics. In almost all cases, the non-significant estimates of treatment effects were in the desired direction but mirrored non-significant pre-intervention differences.

OUTCOMES OF AN ELEMENTARY GRADES SOCIAL COMPETENCE  
EXPERIMENT ACCORDING TO STUDENT SELF-REPORT

By

Elise Touris Harak

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2008

Advisory Committee:  
Professor Gary Gottfredson, Chair  
Associate Professor William Strein  
Professor Sylvia Rosenfield  
Professor Denise Gottfredson  
Professor Dennis Kivlighan

## Acknowledgements

I would like to extend my most sincere thanks and gratitude to the people who have supported me through my doctoral program and dissertation process. I never could have finished without the support, time, and passion given to me by others. I truly appreciate everything you have done. As they say, it takes a village!

Throughout my studies at the University of Maryland, each of the professors instilled in me a passion for the field of school psychology. What you have provided in each class, research project, and mentoring experience have made me a better researcher and practitioner. A piece of each of you is in my work.

Had I not met my dissertation chair and advisor, Dr. Gary Gottfredson, the path I am now on would have been substantially different. Gary, I cannot thank you enough for giving me the opportunity to work on such a fantastic project and for believing in my abilities as a researcher, long before I found my excitement for research. There are no words to adequately express my gratitude.

I would also like to thank my committee members, Dr. William Strein, Dr. Sylvia Rosenfield, Dr. Denise Gottfredson, and Dr. Dennis Kivlighan. Your encouragement, thoughtful feedback, and suggestions have truly made this dissertation complete. It has been a pleasure working with each of you.

This project overall comes from the hard work of my fellow graduate research assistants--Allison Nebbergall and Joe Nese. Particularly once I was working off-campus, you were always quick to send or explain data to me and to talk through aspects of the research. On a personal note, you made working on this project fun and were able to share in my joys and grief about it.

I would also like to thank the University of Maryland “dissertation support group.” Lauren Kaiser, thank you for coming up with the idea of meeting regularly to track research progress. Sharon Huang, Tracey Potter, Sarah Groff, and Sara Signor – thank you for providing me with support and encouragement and for helping me stick to my goals! Sharon, I extend an extra special thank you to you for reading and editing my drafts.

Finally, I must thank my extremely loving and supportive family for encouraging me when I needed it most and for always being so proud of me. Without that pride, there were times I may have given up. This is especially true of my fiancé, Aaron, who has been there through it all. Through your love and support, you have helped me realize my dreams and aspirations and I love you for that!

## Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables.....	v
Literature Review.....	1
The Present Focus.....	2
Prior Related Studies.....	4
Aggression and Problem Behaviors.....	5
Social Skills and Prosocial Behaviors.....	11
Academic Commitment and Achievement.....	15
Differential Effectiveness.....	18
<i>Whom the intervention targets</i> .....	19
<i>Interactions with risk</i> .....	19
<i>Program theory</i> .....	20
<i>Demographic characteristics</i> .....	21
<i>Summary</i> .....	23
Present Study.....	23
Methods.....	24
Participants.....	24
Intervention.....	27
<i>Implementation</i> .....	29
Student Survey Self-Reports.....	31
Variables.....	33
Analyses.....	36
<i>Sensitivity analyses</i> .....	38
Results.....	40
Sensitivity analyses.....	44
Discussion.....	46
Strengths.....	48
Limitations.....	49
Author Note.....	55
Footnotes.....	56
Appendix A.....	57
Appendix B.....	58
Appendix C.....	59
Appendix D.....	63
Appendix E.....	64
References.....	72

List of Tables

Table 1: Students and Data Available in the Study.....27

Table 2: Percentage of Students Receiving Scale Scores on the Student Self-  
Report.....28

Table 3: Second Step Implementation for the 2004-5 and 2005-6 School Years.....31

Table 4: Correlations between Spring 2006 Student Survey Scales.....33

Table 5: Conceptual Descriptions, Transformations, Interpretations, and ICCs for  
Each Scale.....35

Table 6: Second-Year Effect Sizes in Benchmark and Sensitivity Analyses.....43

Table 7: Treatment Coefficients When Predicting Baseline Student Survey  
Scales.....45

Table 8: Second-Year Effect Sizes When Third Graders Were and Were Not  
Included.....46

  

Table E1: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Engagement in Learning Made by Students, Third and Fifth Grade  
Sample.....64

Table E2: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Empathy Made by Students, Third and Fifth Grade Sample.....65

Table E3: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Self-Restraint Made by Students, Third and Fifth Grade Sample.....66

Table E4: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Altruism Made by Students, Third and Fifth Grade Sample.....67

Table E5: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Hostile Attribution Bias Made by Students, Third and Fifth Grade  
Sample.....68

Table E6: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Aggression Made by Students, Third and Fifth Grade Sample.....69

Table E7: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Rebellious Behavior Made by Students, Third and Fifth Grade Sample.....70

Table E8: Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of  
Victimization Made by Students, Third and Fifth Grade Sample.....71

## Literature Review

A constellation of problem behaviors—aggression, classroom misconduct, school failure and dropout, violence, and substance use and abuse—are a concern in schools (G.D. Gottfredson, 1987; Hawkins, Catalano, & Miller, 1992; Moffitt, 1993; Weissberg & Elias, 1993). These behaviors were once seen as separate problems, but these behaviors have common etiologies (Hawkins, Catalano, & Miller, 1992). In general, children are more at-risk to display problem behaviors when they experience harsh and erratic discipline, weak attachment to family members, or have family members or friends who abuse substances or commit crimes (G.D. Gottfredson & D.C. Gottfredson, 1999; Hawkins, Catalano, & Miller, 1992). Individual predispositions that also put children at-risk include impulsivity, irritability, attribution bias, aggressiveness, social isolation, and the rejection of social rules (G.D. Gottfredson, 1987; G.D. Gottfredson & D.C. Gottfredson, 1999; Hawkins, Catalano, & Miller, 1992). Protective factors against problem behaviors include social problem-solving skills, positive relations with adults, attachment to school, commitment to education, and acceptance of conventional social rules and norms (Hawkins, Catalano, & Miller, 1992; Hirschi, 1969).

Because children spend a large part of their time in schools, schools are an ideal place for prevention and intervention related to problem behaviors. Problem behaviors are displayed in schools and often have school-related risk and protective factors (Hawkins, Catalano, & Miller, 1992; S.J. Wilson & Lipsey, 2007). Increasingly, schools explicitly target protective factors through teaching students problem-solving skills, social rules, and ways to make positive choices (D.C.



Gottfredson & G.D. Gottfredson, 2002). This type of social skills curriculum generally focuses on skill building, self-regulation and management, and teaching ways to avoid violence (G.D. Gottfredson & D.C. Gottfredson, 1999). Some schools address the learning environment and student-teacher relations to enhance student attachment to school (Hahn et al., 2007; S.J. Wilson & Lipsey, 2007). An emerging consensus of reviewers implies that prevention programs aimed at increasing protective factors can decrease problem behaviors. These programs may also have positive effects on social skills and academic performance (Hahn et al., 2007; S.J. Wilson & Lipsey, 2007).

Research-supported prevention programs share a number of common features: they (a) use a cognitive-behavioral approach, (b) teach a variety of social and problem-solving skills, (c) are delivered early, preferably in elementary school, (d) are understood and well-implemented by school personnel, and (e) are implemented consistently (Botvin & Kantor, 2000; Conduct Problems Prevention Research Group [CPPRG], 1999; D.C. Gottfredson & G.D. Gottfredson, 2002; Weissberg & Elias, 1993; Wilson & Lipsey, 2007). Implementation of prevention programs in schools is often poor (D.C. Gottfredson & G.D. Gottfredson, 2002). In order to ensure that implementation occurs, there should be: (a) clearly stated goals, (b) leadership support for the activities, (c) adequate training for personnel, and (d) implementation monitoring (G.D. Gottfredson et al., 2000).

### *The Present Focus*

The present research focuses on one plausible program of interest, Second Step (Committee for Children, 1991). This popular program has been endorsed by

many agencies including the U.S. Department of Education Office of Safe and Drug-Free Schools, Office of Juvenile Justice and Delinquency Prevention, Collaborative for Academic, Social, and Emotional Learning, the U.S. Department of Health and Human Services Substance Abuse and Mental Health Services Administration, and the Center for Substance Abuse Prevention (Committee for Children, n.d.). Many of the components of Second Step resemble those found in interventions that have been shown to have efficacy. Second Step emphasizes a blend of research-based techniques from social learning, social information-processing, cognitive-behavioral, and self-regulation theories. It teaches students to: (a) identify social problems, (b) understand others' feelings and intentions, (c) think through non-aggressive or prosocial ways to solve social problems, and (d) evaluate solutions. The assumption is that through modeling, teaching, and feedback about these skills, children will think differently about social situations and behavior will change (Frey et al., 2000; Hahn et al., 2007). These skills directly target risk factors (impulsivity, attribution bias, peer rejection) and potentially effect protective factors (positive relations with adults, attachment to school).

Second Step can be implemented by classroom teachers. It is an easy-to-use, manualized curriculum. To enhance implementation, the Committee for Children provides training to school personnel and supports the school-wide, long-term use of the program (Frey, Hirschstein, & Guzzo, 2000). There are lesson kits for preschool, elementary, and middle school teachers.

### *Prior Related Studies*

The following paragraphs review studies of the efficacy of social skills instructional programs with a focus on those programs and studies with implications for the evaluation of the Second Step program on which the present research will focus. Evaluations of prevention programs have predominantly focused on decreases in aggression. Social skills, commitment to learning, and academic achievement have also been examined. The following sections review effects on aggression, social skills, prosocial behaviors, academic commitment, and achievement according to meta-analyses, exemplary program evaluations, and evaluations of the Second Step program specifically. Some programs reviewed here were categorized as model programs by Hahn et al. (2007). Many of the evaluations included use a suitable research design (i.e. few threats to validity, a control group, or multiple data points; Hahn et al., 2007). Evaluations of model or popular programs were excluded if they did not use a suitable research design. Evaluations of Second Step with weak research designs are reviewed.

The programs examined here seek to enhance elementary students' "protective factors" to decrease the development of problem behaviors. Some programs explicitly teach all students social skills and ways to avoid violence. Other programs target academic commitment and academic success through teacher training in classroom management and instruction or by focusing on positive relations between teachers, parents, and students (Hall & Bacon, 2005; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Jalongo et al., 1999; Olweus, 1997; Reid, Eddy,

Fetrow, & Stoolmiller, 1999; Shure & Spivack, 1982; Stern, 1999; Weissberg et al., 1981).

Individual studies were excluded if they did not assess the outcomes of interest: problem behaviors, social skills, attitudes, and achievement. “Problem behaviors” include verbal and physical aggression, poor classroom conduct, misconduct in school, and delinquency. “Social skills” refer to social problem-solving or conflict resolution techniques, self-control, and prosocial or helping behaviors. “Attitudes” encompass feelings about problem behaviors, perceptions of others’ behavior and intent, and commitment to school. Because Second Step is a universal program, individual evaluations of programs delivered only to a targeted population are excluded from this review. The present study evaluates program outcomes in elementary schools. Therefore, only evaluations at this school-level are included.

Evidence of potential differential effectiveness for program types and student demographics are also reviewed to help determine which, if any, programs work best and for whom. This discussion draws on and extends earlier reviews (D.C. Gottfredson, 2002; Durlak & Wells, 1997; Hahn et al., 2007; Harak, 2006; Wilson & Lipsey, 2007).

### *Aggression and Problem Behaviors*

Meta-analyses and individual studies of prevention programs provide evidence that social skills programs can decrease aggression. This evidence has been consistent across multiple rigorous meta-analyses with most reporting mean effect sizes ( $M_{ES}$ ) ranging from .20 to .35 (CPPRG, 1999; Durlak & Wells, 1997; S.J. Wilson & Lipsey, 2005, 2006, 2007). The 95% confidence intervals for effect

estimates nevertheless sometimes include null and negative effects. One analysis reports an effect size range of  $-.45$  to  $2.36$  (Durlak & Wells, 1997). Hahn et al. (2007) reported that the median reduction for violent behavior was 18% when programs were delivered in elementary schools. Many individual evaluations reported statistically significant effects on aggression.

Evaluations of Second Step specifically do not provide conclusive evidence with respect to aggression. Harak (2006) reported first-year student self-report findings for the present randomized, controlled trial in 12 schools. After one year, fourth- and fifth- graders' self-reports did not show treatment effects on measures of Rebellious Behavior, Aggression, Acceptability of Aggression, Victimization, or Hostile Attribution Bias. Grossman et al. (1997) completed a randomized, controlled trial evaluating Second Step in nineteen second- and third-grade classrooms, across six schools. After one year of implementation, there were decreases in physical aggression on the playground based on researcher observations. Effects were not present on observations of physical aggression in the classroom, verbal aggression in any setting, or on parent and teacher ratings.

Frey, Nolan, Edstrom, and Hirschstein (2005) also studied Second Step claiming that they used a randomized, controlled trial in 15 schools. Eleven schools were randomly assigned in the pre-study year. Two-thirds of these 11 schools were assigned to treatment and one-third to control. After one year of implementation, four more schools were recruited and all were asked to be control schools. After one year, Frey et al. reported that treatment students were rated as significantly less antisocial by teachers. These findings were not replicated in the second year. Observations and

student surveys were also used but outcomes were unclear in the article. Issues of attrition (37% of students did not receive consent and nearly 30% moved after the first year) and low implementation went unaddressed. It is difficult to argue that the Frey et al. study provides support that Second Step is effective at reducing student aggression.

McMahon, Washburn, Felix, Yakin, and Childrey (2000) completed a weak, non-experimental pretest-posttest study of Second Step. After one year of implementation, there were posttest decreases on teacher ratings of problem behaviors for pre-k students. Students in kindergarten showed increased problem behaviors at posttest on the teacher ratings. In contrast, observations revealed a greater decrease in problem behaviors for kindergartners as compared to preschoolers. Cooke, Ford, Levine, Bourke, Newell, and Lapidus (2007) also completed a non-experimental pretest-posttest study of Second Step in five elementary schools. After one year of implementation, students reported significant decreases in impulse control and increases in aggressive behaviors at posttest. On observations and discipline referrals, too few incidents were present to compare pretest to posttest performance. In two additional non-experimental, unpublished dissertations, significant decreases in aggression were found for one grade within the sample, but not in the other grades examined (Botzer, 2002; Riese, 2004). In short, prior studies provide no consistent, credible evidence that Second Step affects aggression; most evaluations have weak research designs.

Four programs that provide direct instruction in problem-solving and social competence skills, like Second Step, found positive effects on aggression and other

problem behaviors. The Too Good for Violence program focuses on problem-solving skills, anger management, and effective communication. Hall and Bacon (2005) evaluated this program using a randomized, controlled trial. After five months of implementation, there were statistically significant effects on teacher ratings of student behavior (Hall & Bacon, 2005). The What Works Clearinghouse (WWC) estimated the effect size as .50 on student behavior (Institute of Education Sciences, n.d., p. 3, 5).

Weissberg et al. (1981) completed a quasi-experimental evaluation of the Social Problem-Solving Skills training in eight schools. Teachers chose whether they were a treatment or control classroom. Lessons were delivered by teachers to second through fourth graders three times per week for 14 weeks. The researchers reported an effect size of .26 on teacher ratings of problem behaviors (Weissberg et al., 1981). Other evaluations of the Social Problem-Solving Skills training had weak designs or were not in elementary schools and therefore are not reviewed.

The Interpersonal Cognitive Problem Solving (ICPS) program was evaluated by Shure and Spivack (1982) using a quasi-experimental-with-comparison-group design. Lessons were delivered daily for eight weeks to Black, inner-city students when they were in pre-k and kindergarten. At posttest, there was a significant difference between the number of treatment and control students who were rated as “adjusted” (Shure & Spivack, 1982, p.345). This difference was not present at baseline. A rating of “adjusted” was based on teacher ratings of student interpersonal behaviors and on the student’s ability to generate multiple solutions to problems and name consequences for behaviors (Shure & Spivack, 1982). This study had the

strongest research design amongst published ICPS evaluations completed in elementary schools.

The Unique Minds School Program (Stern, 1999) targets an increase in student self-efficacy, problem-solving, and social competence and the creation of a positive classroom climate. The goal is to prevent problem behaviors and enhance academic achievement. Linares et al. (2005) evaluated this program using a quasi-experimental-with-comparison-group design. After two years of implementation, there was an effect size of .51 on teacher ratings of aggression (Linares et al., 2005).

Three programs delivered social skills training to students as one component of the intervention. They all report some effects on aggression. Reid et al. (1999) completed a randomized, controlled trial in 12 elementary schools, evaluating the Linking the Interests of Families and Teachers (LIFT) program. The social skills intervention was delivered to first and fifth graders by school and research staff. Another component of this project was parent involvement. After 10 weeks, there was an effect size of -.11 on observations of physical aggression on the playground (Reid et al., 1999). There was also an interaction with baseline aggression for first graders. Effects of the intervention were larger for first graders who were more aggressive at baseline.

The Seattle Social Development Project included cognitive social skills training and teacher training in instruction and classroom management. Hawkins et al. (1999) evaluated this project six years after the intervention was delivered using a quasi-experimental-with-comparison-group design. The intervention was



implemented when the students were 12. At the age of 18, fewer treatment students reported engaging in violence and school misbehavior than the control students.

The classroom-centered intervention targets academics and behavioral management, and uses the Good Behavior Game. Ialongo et al. (1999) evaluated this intervention using a randomized block design. Three classrooms in nine schools each were randomly assigned to treatment or control. After two years, teachers rated treatment students as displaying significantly fewer problem behaviors than the control students. The effect sizes were reported separately for boys and girls for both years. Effect sizes ranged from .25 to .49. There was also an interaction effect with implementation fidelity. Effects of the classroom-centered intervention were highest for the classrooms that implemented at least 50% of the intervention (Ialongo et al., 1999).

One final program did not include a skill training component. The Bullying Program was evaluated by Olweus (1997) using a controlled trial in 42 Norwegian schools. This intervention calls adults' attention to bullying behavior and defines it. After eight months, Olweus found large effect sizes ( $d > 1.00$ ) for aggressive bullying, victimization, and delinquent acts according to student self-reports.

Of the evaluations reviewed, aggression often was the main outcome of interest. Aggression is especially important because it tends to be a stable trait; young children who are aggressive tend to be aggressive in later adolescence (G.D. Gottfredson, 1987; Hawkins, Catalano, & Miller, 1992; Moffitt, 1993). The evaluations of Second Step specifically do not provide support that this program decreases aggression. Support for the hypothesis that social skills programs can

decrease aggression is provided through meta-analyses and the evaluations of other programs.

### *Social Skills and Prosocial Behaviors*

Meta-analyses and individual program evaluations less frequently focus on social skills and prosocial behaviors than on aggression. This is a weakness in the literature. Many prevention programs directly teach social problem-solving or other social skills. Evaluations including social skill and prosocial behavioral outcomes would be helpful in examining their relation to problem behavior.

Social skills are measured in two ways; treatment-inherent measures that examine specific knowledge or skills taught by the curriculum and more distal measures where social competence behaviors are observed and rated. The effect sizes differ for these two kinds of measures, with treatment-inherent measures showing much larger effects than the more general, non-treatment-specific, measures. Large effects on treatment-inherent measures indicate that the students have learned new concepts and can verbalize what skills they should use. This does not equate with application of the skills in everyday life. Only some studies have demonstrated improved behaviors or use of skills (Flannery et al., 2003; Grossman et al., 1997; Linares et al., 2005; Reid et al., 1999)

S.J. Wilson and Lipsey reported effect sizes between .30 and .40 for social skills in their 2005, 2006, and 2007 meta-analyses. These effect sizes include effects on both treatment-inherent and distal measures. Other meta-analysts did not include social skills or prosocial behaviors as an outcome. When individual studies evaluated the effects of prevention programs on social skills, some found positive effects using

a variety of outcome measures. These effect sizes range from .14 on teacher ratings of skills (Reid et al., 1999) to 1.01 on tests of content knowledge (Linares et al., 2005).

One experimental evaluation of Second Step provides limited support that it is effective at increasing students' observable social skills. A randomized trial by Grossman et al. (1997) found that after one year, treatment students showed a statistically significant increase in prosocial and neutral behaviors via observations on the playground and in the cafeteria. Treatment effects were not detected using parent or teacher ratings.

In the first-year experimental evaluation of Second Step, Harak (2006) measured effects using students' self-reports in the present sample. There were no treatment effects on measures of Altruism, Empathy, or Self-Restraint. Interaction effects with individual characteristics were also tested. There was an interaction of treatment with students' baseline score on the Self-Restraint scale. Treatment students with below average self-restraint at baseline reported statistically significant increases in their self-restraint after one year of implementation.

Frey et al. (2005) also studied Second Step. The authors claimed that there was random assignment, though four control schools were not randomly assigned. After one year of implementation, treatment students were rated by teachers as displaying significantly more social competence. These findings were not replicated in the second year. The findings from other sources of data (observations and student surveys) were so unclear that it is difficult to draw any conclusions from this study.

McMahon et al. (2000) used a non-experimental, pretest-posttest design to evaluate Second Step for one year. They reported that students exhibited more

knowledge and skills related to the curriculum on knowledge tests at posttest. Effects were more pronounced in the kindergarten children than in the pre-k children. Cooke et al. (2007) also completed a non-experimental pretest-posttest study of Second Step in five elementary schools. After one year of implementation, student-self reports reflected increases in positive coping skills and cooperative behaviors. On observations, fewer prosocial behaviors were seen at posttest than at pretest. The weak designs of these two studies, the limited scope, and mixed outcomes provide little evidence that Second Step improves students' social skills or behaviors. Other studies evaluating Second Step did not find significant effects on prosocial behaviors or social skills (Botzer, 2002; Grossman, et al., 1997; Harris, 1998; Lillenstein, 2001; McDonald, 2001; McMahon et al., 2004). For a full review, see Harak (2006).

Other programs that affected social skills focused on teaching problem-solving techniques similar to Second Step. The following paragraphs review effects on social skills for these other programs.

The experimental evaluation of Linking the Interests of Families and Teachers (LIFT) showed effects on teacher ratings of social skills after one year of implementation ( $ES = .14$ ; Reid et al., 1999). LIFT taught skills but also targeted parental involvement.

In the two-year, quasi-experimental evaluation of the Unique Minds School Program, there were effects on three social skills measures. Treatment students reported feeling more socially efficacious ( $ES = .55$ ) than did control students. Students also displayed more knowledge of prosocial problem-solving strategies

according to interviews with observers ( $ES = 1.01$ ). Finally, teachers rated treatment students as more socially competent ( $ES = .48$ ; Linares et al., 2005).

A quasi-experimental evaluation of the Social Problem-Solving Skills training also demonstrated significant differences between treatment and control students after 14 weeks of implementation (Weissberg et al., 1981). These effects were seen on a student interview: treatment students responded with more solutions for social problems ( $ES = .49$ ), shared more effective solutions, and could better verbalize the problem-solving steps. The effect size was .88 for the total score on this interview (Weissberg et al., 1981).

A quasi-experimental evaluation of the Interpersonal Cognitive Problem Solving (ICPS) program also included an interview outcome measure. After eight weeks of implementation, treatment students showed statistically significant gains in problem-solving skills (Shure & Spivack, 1982).

An experimental evaluation of the Peace Builders program was completed by Flannery et al. (2003). After a year-long implementation, teachers rated the treatment students as more socially competent than they did the control group. During the second year, both groups were given the treatment. At the end the second year, the initial treatment group was still rated as more competent (Flannery et al., 2003).

Evaluations through student self-report for two programs did not find statistically significant effects on social skills. Neither the experimental evaluation of Too Good for Violence by Hall and Bacon (2005) nor the quasi-experimental evaluation of the Heartwood Ethics Curriculum done by Leming (2000) found effects on social skills. The authors of both studies cited ceiling effects as a possible cause. It

could also be that student self-report measures are less sensitive to change than interviews, observations, and teacher ratings.

Almost all of the studies reviewed included a skills training component. Despite this, many do not examine effects on prosocial behaviors or social skills. Of the studies reviewed, many excluded prosocial behaviors as outcomes. When social skills are included as an outcome, researchers often report significant effects. Effects are more pronounced when treatment-inherent measures are used. These measures may not translate into observable behavior change.

#### *Academic Commitment and Achievement*

Evaluations of elementary school programs emphasizing protective factors against problem behaviors often exclude academic commitment and achievement as outcomes. These outcomes are prevalent in other fields of study. Commitment or attachment to school is important because there is a strong positive correlation between commitment and social development and academic achievement (G.D. Gottfredson, 1999; G.D. Gottfredson & D.C. Gottfredson, 1999; Hirschi, 1969). Students who feel a bond or commitment to school are less likely to experience school failure or engage in behaviors such as aggression, classroom misconduct, violence, or substance abuse (Gottfredson, 1987; Hawkins, Catalano, & Miller, 1992; Hawkins et al., 1999; Moffitt, 1993; Weissberg & Elias, 1993).

S. J. Wilson and Lipsey (2005) reviewed 82 studies examining classroom participation and 148 examining academic achievement outcomes. They reported mean effect sizes of .22 and .19, respectively. Their updated meta-analyses in 2006 and 2007 reported nearly identical effects. In all cases, confidence intervals did not

overlap with zero. Hahn et al. (2007) cited four studies where “substantial improvements in school attendance and achievement” (p.123) were noted.

The evaluations of programs that only taught social skills rarely included academic commitment, participation, or achievement outcomes (Grossman et al., 1997; Hall & Bacon, 2005, Shure & Spivack, 1982; Weissberg et al., 1981). Most studies evaluating Second Step did not assess academic commitment or performance. In the first-year student self-report evaluation of Second Step in the present sample, a main effect for treatment was found on student self-reports of engagement in learning after one year of implementation ( $ES=.11$ , Harak, 2006). Students in treatment schools reported being more attentive and involved in their classes than control students. This engagement measure is similar to what Hirschi (1969) calls commitment to education.

Using a posttest-only design with a non-equivalent comparison group to evaluate Second Step, McDonald (2001) reported higher math and reading achievement scores for the comparison group than the treatment group. McDonald’s study cannot be interpreted because of its weak design.

Elias et al. (1991) used a posttest-only design to evaluate the Improving Social Awareness-Social Problem Solving Project (ISA-SPS) six years after it was implemented. The treatment group was divided into two categories; high and low fidelity. Only the high fidelity group showed significantly higher standardized scores on language arts and math at follow-up.

The evaluations reviewed in the following paragraphs include a wide range of interventions targeting school and classroom climate changes, teacher training,

improved behavior management, and academic interventions. Few studies examined academic commitment. Overall, the studies do not provide conclusive evidence of effects on achievement.

In an experiment evaluating the effects of the classroom-centered intervention on achievement, Ialongo et al. (1999) did not report any main effects on achievement, but they did report interaction effects for boys only. On reading achievement, there were effects for boys at or above the 40<sup>th</sup> percentile at baseline. On math achievement, there were effects for second grade boys who achieved at or above the 60<sup>th</sup> percentile at baseline.

In the quasi-experimental evaluation of the Seattle Social Development Project, Hawkins et al. (1999) found that treatment students reported significantly higher attachment, commitment to school, and achievement than control students, six years after the intervention was implemented. Treatment students also showed near-significant gains on their high school grade point average and decreases in dropout rates. There were no significant differences between the groups on standardized achievement test scores.

The two-year, quasi-experimental evaluation of the Unique Minds School Program showed statistically significant improvements on math report-card grades ( $ES = .42$ ). The effects on reading report-card grades were not significant ( $ES = .24$ ; Linares et al., 2005). In contrast, there was a negative, but non-significant effect on math standardized test scores ( $ES = -.19$ ) and no effect on reading scores (Linares et al., 2005).



In a quasi-experimental evaluation, Flay and Allred (2003) reported statistically significant effects of the Positive Action intervention on a Florida Reading Test, but not on the Florida Comprehensive Aptitude Test (FCAT, tests given only to fourth graders). This finding should be interpreted with caution. It appears that there were significant baseline differences, favoring the treatment schools, on reading and writing. The authors do not report controlling for pretest differences. The What Works Clearinghouse estimated that the effect size on the Florida Reading Test was 1.29 (Institute of Education Sciences, n.d., p. 8).

G.D. Gottfredson, Jones, and Gore (2002) examined the Social Problem-Solving curriculum using a non-equivalent comparison group design with pre-intervention covariates. They did not detect any effects on academic commitment or achievement outcomes. This likely is due to poor program implementation.

In summary, few evaluations of programs similar to Second Step examine academic commitment and achievement. The findings do not provide consistent evidence that social skills curricula affect academic commitment or achievement. Evaluations of programs that address positive climate, teacher training, and classroom and behavior management more often include these outcomes. One study provides some evidence that academic commitment is positively affected after implementation (Harak, 2006). Few show that objective measures of achievement (standardized assessments) are affected.

#### *Differential Effectiveness*

A major focus of the meta-analyses on social skills programs is differential effectiveness based on the type of program used. Analysts have examined differences

in how the interventions are delivered (universally, to selected or to indicated samples), and they have classified studies in terms of theoretical underpinnings of the programs. Researchers have also examined how interventions interact with personal characteristics of the sample. The following paragraphs discuss these program features in light of results from meta-analyses. Second Step is a universal program, delivered through social vignettes and classroom discussion, using a cognitive-behavioral approach.

*Whom the intervention targets.* According to S.J. Wilson and Lipsey (2005), universal programs (those delivered to everyone) are less effective ( $M_{ES} = .18$ ) than those which also have a selected or indicated component ( $M_{ES} = .30$  for both). In 2006 and 2007, S.J. Wilson and Lipsey reported approximately the same effects as in 2005. Hahn et al. (2007) reported that across all grade levels, universal programs resulted in a 15% reduction in violent behavior. One possible reason for the difference in effectiveness between universal and targeted programs is the difference in how these programs are studied. S.J. Wilson and Lipsey (2006) noted that most studies of targeted programs were randomized, whereas most universal program evaluations were not.

*Interactions with risk.* When meta-analyses compared programs targeting youths at-risk of, or already exhibiting, behavioral problems, effects were larger than when only the general population was targeted (S.J. Wilson & Lipsey, 2005, 2007). This is supported by findings by Beelman et al. (1994) that effects for students who were socially deprived or have been affected by adverse life events ( $M_{ES} = 1.06$ ) were larger than for students who: (a) were not socially deprived, (b) had not experienced

adverse life events, (c) did not display externalizing or internalizing symptoms, or (d) did not have learning problems ( $M_{ES} = .75$ ). Interestingly, the effects for students with externalizing or internalizing symptoms were the smallest ( $M_{ES} = .67$ ; Beelman et al., 1994).

Another way to examine interactions with risk is by testing for interactions between baseline performance and treatment status within individual program evaluations. Flannery et al. (2003) reported that in their year-long, experimental evaluation of the Peace Builders Program, students who performed poorly at baseline showed the greatest positive change. A significant treatment-by-initial-status interaction was not discussed (Flannery et al., 2003). S.J. Wilson and Lipsey (2007) suggested that interactions with baseline occur because students who already exhibit problems have more room to improve.

*Program theory.* The theoretical frameworks of interest include: (a) behavioral social skills training, (b) cognitive social skills training, and (c) cognitive-behavioral social skills training. There is evidence that when comparing these types of programs, there is no effect for the approach on the reduction in violence or aggression (Hahn et al., 2007, S.J. Wilson & Lipsey, 2005, 2006, 2007). In other words, all of these types of programs result in a reduction of aggression. In general, all three types of programs result in small to medium effects, with effect sizes ranging from .10 to .43. (D.B. Wilson et al., 2001, S.J. Wilson & Lipsey, 2003, 2005). The nature of the assessment used to measure outcomes seems more important than program type in determining the size of the effect.

*Demographic characteristics.* The sample demographics reviewed are baseline performance, age, socioeconomic status, gender, and ethnicity. Very few studies examined or found differential effectiveness on gender and ethnicity.

*Gender.* One individual experimental evaluation found achievement effects for boys but not girls. On all other outcomes in this evaluation, effects were found for both genders but varied in size (Ialongo et al., 1999).

*Ethnicity.* One meta-analysis examined whether the effects of prevention programs varied based on the predominant ethnicity in a school building (Hahn et al., 2007). Hahn et al. (2007) reported that the reduction of violence in schools where more than 50% of students were Black was the same as in schools where more than 50% of students were White.

*Age.* Effectiveness differs by age group or school level (preschool, elementary, middle, and high). For universal programs targeting a reduction in aggression and violence, the largest median reduction was 32% in pre-k and k programs followed by high school (29% reduction), elementary school (18%), and then middle school (7%) (Hahn et al., 2007). S.J. Wilson and Lipsey (2007) reported that effects of universal programs were largest for the youngest students. With the exception of preschool-aged children, the lower bound of effectiveness of affective education and interpersonal programs is similar for all age groups; the differences lie in the upper bound estimates of effects (Durlak & Wells, 1997).

*Socioeconomic status.* Universal programs targeting a reduction in aggression or violence are often used in schools where the neighborhood has higher crime and lower socioeconomic status (SES; S.J. Wilson & Lipsey, 2006, 2007). According to

Hahn et al. (2007), universal programs are as effective in these neighborhoods as they are when crime and low SES are not present. S.J. Wilson and Lipsey (2006) presented conflicting evidence. They found an interaction between age and SES. Younger students who came from families of low SES benefited most from prevention programs. It is unclear whether SES matters for the effectiveness of prevention programs.

*Research design.* The evidence is mixed for the correlation between better designs and the likelihood of detecting effects. S.J. Wilson and Lipsey (2005) reported that factors relating to the research design accounted for 36% of the variance in effect sizes. When randomization was used, the mean effect size was .35. Effect sizes were substantially smaller in studies that did not randomize. In a test of heterogeneity of findings, S.J. Wilson and Lipsey (2007) reported significant variability in effect sizes across studies reviewed. S.J. Wilson and Lipsey (2007) and Hahn et al. (2007) did not find a significant effect of design on outcomes, however. Effect estimates for prevention programs remained the same whether studies with the least suitable designs were included or excluded (Hahn et al., 2007).

Using measures of implementation integrity, multiple archival and observation outcome measures, and teacher reports are all associated with larger effect sizes (S.J. Wilson & Lipsey, 2007). Studies with high rates of attrition are less likely to detect effects, but this association was not always significant (S.J. Wilson & Lipsey, 2005, 2006; S.J. Wilson, Lipsey, & Derzon 2003).

There are arguments in the literature that students need a sufficient exposure to treatments, but Hahn et al. (2007) found that there was no clear connection

between larger effects and either the number of sessions per week or the number of months that a program was implemented. S.J. Wilson and Lipsey (2006) found that implementation integrity did not significantly predict effect size, but the relationship was in the expected direction (better implementation is associated with larger effects). S.J. Wilson and Lipsey also found that programs implemented over a longer period of time were less effective. This could be because long-term programs are less intense than short-term programs.

*Summary.* For a universal, cognitive-behavioral program delivered to elementary-aged students, the literature suggests that effects of about .20, or two-tenths of a standard deviation, can be expected on aggression, academic commitment, and achievement. Slightly larger effects (.30-.40) would be expected on social skills. It is possible that these programs have the best results when delivered in schools with higher need and to students who already show problem behaviors. When using randomization and ensuring high implementation, it is possible that effects will be larger.

#### *Present Study*

Programs targeting the prevention of problem behaviors have often produced evidence of effectiveness in decreasing problem behaviors and increasing social competence. There is less evidence supporting the effects of prevention programs on academic commitment and achievement. Evaluations of Second Step have not generally provided this kind of support. There are only two experimental evaluations and only a couple of non-replicated, significant effects. The present study is an evaluation of the effectiveness of Second Step when implemented for a two-year

period as measured by student self-report. This study is an extension of Harak (2006) which evaluated first-year findings on eleven student survey scales.

In considering the findings of prior research on Second Step and similar programs, the research questions for the present study are:

1. How do two years of the Second Step curriculum affect students' self-reports of their own engagement in learning, social competence (empathy, altruism, and self-restraint), and problem behaviors (perception of and engagement in aggressive acts, rebellious behavior, hostile attribution bias, and victimization)?
2. Is there an interaction with free or reduced meals (FARM) status or baseline ratings?

Specific hypotheses are that:

1. Student surveys will reflect a continued effect on Engagement in Learning and an emerging effect on both social competence and problem behaviors.
2. There will be treatment interactions with FARM status and baseline ratings. Students who received free or reduced meals and poor baseline ratings (low grades, high levels of problem behaviors) will benefit most from the Second Step curriculum.

## Methods

### *Participants*

This study is an evaluation of the effects of two years of exposure to Second Step, and it is part of a larger multi-site matched and randomized, controlled trial. The first through fifth grade students and teachers in 12 Anne Arundel County (Maryland) elementary schools were included in the evaluation. The population density and

income of residents varies across this county, but the majority of the schools in this sample however were in affluent, suburban neighborhoods. The schools involved in this study had never implemented Second Step and agreed to be randomly assigned to treatment status. Matched pairs were formed based on school size, ethnic/racial composition, and participation rates in the free and reduced meals (FARM) program. Two schools could not be well matched. One school had 600 students, a size far larger than the next largest school. Therefore, the matched school had about 125 fewer students. A second school had a population 75% composed of ethnic minorities. The school with the next largest population of ethnic minorities had about 30%. Thus, the matching of the schools was not exact. Using computer-generated pseudo-random numbers, one member of each pair was randomly assigned as treatment and as a control.

The present student self-report study is limited to students in the third and fifth grades during the 2005-6 school year. Fourth grade students during the 2005-6 school year were part of the national cohort. Data were collected by a contractor on behalf of the Institute of Education Sciences (IES) for the fourth graders. Reports on data collected for that cohort are embargoed pending the release of a report by IES. First and second graders were too young to complete a student survey. Because the 2005-6 third graders were in the second grade when baseline student surveys were administered, no baseline is available. Table 1 displays the data collected at each grade level across two years.



Table 1

*Students and Data Available in the Study*

Grade level and data	School Year	
	2004-5	2005-6
Grade 3-parent, teacher, student	Not available*	Available
Grade 4-parent, teacher, student	Available	Not available*
Grade 5-parent, teacher, student	Available	Available

\*Embargoed pending report by Institute of Education Sciences.

At the beginning of the 2005-6 school year, there were 1,872 students and 78 teachers in grades three and five, across the 12 schools. By the end of the school year, the sample comprised 1,876 students. Of these, 1,354 students received parental consent to participate. Only the 1,314 students in the third and fifth grade in 2005-6, who completed the 2006 student self-report survey, are the focus of this study. Of the students who completed surveys, 86.4% of students are White, 8.1% are African American, 2.7% are Asian, 2.6% are Hispanic, and .3% are American Indian. This ethnic breakdown is similar to the whole sample of third and fifth graders (85.4% White, 8.6% African American, 3% Asian, 2.8% Hispanic, and .3% American Indian). Only 3% of students qualified for reduced lunch and 7% qualified for free lunch in the whole sample of third and fifth graders.

Due to item non-response, the sample size for both grades varies across scales ( $n = 559-1312$ ). When more than half of the items in a scale were left blank, a scale score was not calculated. The five scales that appear earliest in the survey (Engagement, Altruism, Empathy, Self-Restraint, and Hostile Attribution Bias) have

over 1,000 respondents. The three scales that ask about rebellious and violent behavior have about 600 respondents. Though falling later in the survey than the scales about aggression and rebellion, the victimization scale has 920 respondents. Thus, it appears length of the survey and sensitivity of the scales played a role in student non-response.

Table 2

*Percentage of Students With Scores on the Student Self-Report*

Scale	3rd Graders	5th Graders	Total
Engagement in Learning	69	71	70
Empathy	69	71	70
Self-Restraint	67	70	69
Altruism	68	71	69
Hostile Attribution Bias	67	70	69
Aggression	31	39	35
Rebellious Behavior	29	33	31
Victimization	48	50	49
Acceptability of Aggression	28	31	30

*Note.* Percentages were computed by dividing the number of scale scores present by total enrollment. Percentages reflect both the consent rate and scale response rates.

Table 2 displays the overall percentage of students in each grade who responded to each survey scale. Despite concerns that the survey may have been too long for 3<sup>rd</sup> graders, they completed items at a comparable rate to the 5<sup>th</sup> graders. The 3<sup>rd</sup> grade students appeared more sensitive to the negative behavior scales, particularly Aggression. Further analyses exploring the non-response to survey scales are presented in the results section.

*Intervention*

Students participated in Second Step for two years. Second Step is a universal, grade-specific, manualized classroom curriculum that teaches students four areas of

social competence: empathy, problem-solving, impulse control, and anger management. Each of these areas involves several different lessons. Grade five has 22 lessons. Grade three has 15 lessons. Second Step was implemented at all grade levels, but the present research focuses only on students old enough to provide self-reports.

Classroom teachers were provided with their own kit of published materials and delivered one lesson per week. Lessons took between 35 to 45 minutes to deliver and were usually completed on one day per week. All schools were given a schedule for the delivery of each unit, to promote completion of all lessons. The pace at which the schools completed the curriculum varied however with some finishing ahead of schedule and others lagging behind. Pace depended on how early in the school year implementation was started and whether teachers implemented every week. After completing each lesson, teachers were expected to complete a log to indicate which aspects of the lesson were implemented (See Appendix A for a sample implementation log).

The school counselors played an integral role in the intervention and its evaluation, as they served as the liaison between the research team and the teachers in each school. Each month, counselors from all schools met with the research team to discuss issues related to the evaluation. The treatment counselors also received and transmitted to each school implementation feedback data. This was done in order to cope with the difficulties documented in implementing similar prevention programs (D.C. Gottfredson & G.D. Gottfredson, 2002). The counselors helped teachers overcome obstacles revealed by the implementation feedback. This activity was integrated into the treatment schools' "character development team" agenda.

Character development teams existed in treatment schools for the purpose of initiating and monitoring school-wide character education goals and programs, including Second Step. The treatment school counselors were provided training on how to apply the outcome data from this project to school plans and goals. (See Silverman, 2005; or Silverman & Gottfredson, 2005 for further information on implementation feedback.) Control schools engaged in some character education activities, but control school counselors were not provided with guidance from the research team about character education.

*Implementation.* At the end of the 2005-6 school year, all treatment schools reached the implementation standard of completing 80% of the Second Step lessons based on self-report, contemporaneous logs. One school fell short of implementing 80% of the content within the lessons. These standards were set by the research team in collaboration with school personnel prior to program implementation. Within schools, some teachers did not meet the standards. Table 3 displays implementation rates during the first (2004-5) and second (2005-6) year of the study.

An ANOVA showed that there were no significant differences among schools on the percentage of logs submitted ( $F = .79, p = .65$ ) or the percentage of curriculum delivered ( $F = .90, p = .60$ ). There were also no significant differences among grades on the percentage of logs submitted ( $F = 1.13, p = .40$ ) or percentage of curriculum delivered ( $F = 1.43, p = .30$ ). Overall, implementation in the second-year was high (See Appendix B for the first-year findings).

Table 3

*Second Step Implementation for the 2004-5 and 2005-6 School Years*

School Number	Percentage of Logs Completed		Percentage of Curriculum Delivered	
	2004-5	2005-6	2004-5	2005-6
1	100	98	96	89
2	96	99	76	96
3	73	100	63	84
4	93	93	79	79
5	99	100	92	94
6	86	88	76	81

*Note.* The percentage of logs completed is the percentage of logs that were sent to the research team. Key elements of the program were identified (use of videos, distribution of handouts, teacher modeling of skills, student evaluation of model, student role-playing, and generation of problem-solving steps). The percentage of curriculum delivered is the total percentage of implementation of these six elements.

A supplemental teacher survey was administered at the end of the 2006-7 school year to test for differences in the social skills instruction at treatment and control schools. Teachers in treatment schools reported having detailed teacher guides and implementation materials (characteristics of Second Step). Control schools reported having essentially no teacher guides or materials to implement social skills instruction. When tested statistically, there were significant differences between the treatment and control schools for materials and activities (Gottfredson, Harak, Nebbergall, Nese, & Shaw, 2008).

The frequency of social skills instruction also differed. In control schools, 75% of teachers reported implementing social skills instruction twice a month or less frequently. In treatment schools, 81% of teachers reported implementing social skills instruction once per week or more frequently (Gottfredson et al., 2008). This

confirms that implementing Second Step provided treatment schools with a formal, frequently implemented instruction program. Control schools, on the other hand, did not have a comparable program or set of activities in place.

### *Student Survey Self-Reports*

The student self-report survey is composed of 83, four-point Likert-type questions forming 11 composite scales (Harak, 2006). Only nine of the 11 scales are of interest for this study. The two scales evaluating students' perceptions of the school climate are excluded (Sense of School as a Community and Feelings of Safety at School). Student surveys were administered in the fall of 2004 prior to the intervention, in the spring of 2005 after one school-year of implementation, and again in the spring of 2006 after two years of implementation. At each data point, surveys were returned by almost all (97% or more) students who had parental consent.

The student survey was designed by the Institute of Education Sciences (IES) staff in collaboration with the seven principal investigators from the national study.<sup>1</sup> Twelve scales were developed, based on measures used in previous research. The survey was piloted prior to baseline data collection and revised by a data collection contractor prior to use. This survey was created for use with the national cohort.

Our Anne Arundel County project excluded one scale used in the national study, added some questions to the Hostile Attribution Bias scale, and made minor wording changes to some questions for the purpose of this research (Harak, 2006). The survey used with the third and fifth graders in the present study is different in these ways than the survey used with the 2005-6 fourth graders (national cohort).

The student self-report scales were chosen because they measure a variety of risk factors for problem behavior. Also, they tap distal outcomes of thoughts and behavior, rather than being closely curriculum-linked measures. Often, attitudes are unknown and behaviors (positive and negative) can go unnoticed by parents or teachers. Students may provide different insight than others can. According to Wilson and Lipsey (2007), however, effect sizes detected by self-report measures are typically small.

Student survey scales have statistically significant correlations with each other in almost all cases. Correlations do not indicate redundancy however as almost all correlations are smaller than .40. Table 4 displays the scales correlations.

Table 4

*Correlations between Spring 2006 Student Survey Scales*

Scales	1	2	3	4	5	6	7	8	9
Self-Restraint (1)	<b>1</b> (1290)								
Aggression (2)	0.35*	<b>1</b> (664)							
Rebellious Behavior (3)	0.24*	0.62*	<b>1</b> (584)						
Victimization (4)	0.20*	0.32*	0.32*	<b>1</b> (920)					
Acceptability of Aggression (5)	0.17*	0.35*	0.43*	0.14*	<b>1</b> (559)				
Empathy (6)	0.58*	0.23*	0.19*	0.08*	0.17*	<b>1</b> (1307)			
Engagement in Learning (7)	0.46*	0.35*	0.24*	0.21*	0.10*	0.36*	<b>1</b> (1312)		
Hostile Attribution Bias (8)	0.38*	0.28*	0.25*	0.36*	0.13*	0.29*	0.28*	<b>1</b> (1293)	
Altruism (9)	-0.30*	-0.11*	-0.04	0.18*	-0.09*	-0.32*	-0.21*	-0.05	<b>1</b> (1301)

*Note.* Numbers in parentheses beneath correlations indicate the *n* for each correlation.

\**p* < .05.

Scale alphas and stability coefficients (20-month test-retest correlation) were computed with data from this sample and range from .73 to .88 and .18 to .40, respectively. The stability coefficients were computed using fall 2004 and spring 2006 data and are expected not to be high correlations, as that would indicate that the measures are not sensitive to change. See Appendix C for full reliability and stability findings.

#### *Parent and Teacher Surveys*

The parent and teacher survey is a modified version of the Social Competency Rating Form (SCRF; Gottfredson, Jones, & Gore, 2002). Questions were tailored to target an elementary-aged population and skills specific to Second Step (Nebbergall, 2007). The modified SCRF is composed of 29, four-point Likert-type questions assessing problem behaviors (e.g., hits, kicks, or jumps on others) and social skills (e.g., shows self-control when provoked by others). Item responses range from “almost never” (score of 1) to “very often” (score of 4). The SCRF is intended to evaluate intervention programs and is used in our larger study as an outcome measure.

#### *Variables*

Individual students’ scale scores from the spring 2006 student survey assessment are used as dependent variables. The spring 2006 survey scores (used as dependent variables) were skewed in almost all cases. Table 5 lists all of the scales, their conceptual descriptions, the transformations that were made to normalize the data, interpretation of the resulting values, and the proportion of variance between schools (intraclass correlation, ICC). Because positive behavior scale scores were



reflected, higher scores indicated lower levels of positive behavior. Effect sizes were multiplied by -1 so that positive effects indicated increases in positive behavior.

Table 5

*Conceptual Descriptions, Transformations, Interpretations, and ICCs for Each Scale*

Scale Name	Conceptual Description	Transformations and Interpretations	ICC
Engagement in Learning	Assesses how well students attend and participate in class.	The raw data was reflected (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased. Positive effect sizes indicate increased engagement.	.03
Empathy	Assesses social understanding, sensitivity to the feelings of others, and anticipation of the effects one's actions have on others.	The raw data was reflected (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was brought close to zero. Positive effect sizes indicate increased empathy.	.04
Self-Restraint	Measures the extent to which students report that they have the skills to control themselves in challenging social situations.	The raw data was reflected (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased. Positive effect sizes indicate increased restraint.	.02
Altruism	Assesses the extent to which students help others who are hurt or being picked on, cheer others up, or intervene to stop others' aggression.	The raw data was standardized using a z-score transformation (no nonlinear transformation needed). Positive effect sizes indicate increased altruism.	.03
Hostile Attribution Bias	Assesses the tendency to perceive hostile or unfriendly intentions in ambiguous social situations.	The raw data was standardized using a z-score transformation (no nonlinear transformation needed). Positive effect sizes indicate increased attribution bias.	.06

Scale Name	Conceptual Description	Transformations and Interpretations	ICC
Aggression	Assesses physical, verbal, or social aggression against other students.	To normalize, the raw data was transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased. Negative effect sizes indicate decreased aggression.	.07
Rebellious Behavior	Measures the frequency with which students engage in problem behavior such as vandalism, theft, cheating, and skipping school or classes.	The raw data was transformed using the natural log, and standardized using a z-score transformation. Skewness was brought close to zero. Negative effect sizes indicate decreased rebellious behavior.	.06
Victimization	Indicates extent to which students are victims of teasing, name calling, threats, or physical aggression.	The raw data was transformed using the natural log, and standardized using a z-score transformation. Skewness was brought close to zero. Negative effect sizes indicate decreased victimization.	.03
Acceptability of Aggression	This gauges norms of interpersonal aggression such as yelling at or making fun of others and saying bad things.	The raw data was transformed using the natural log, and standardized using a z-score transformation. Skewness was brought close to zero. Negative effect sizes indicate decreased acceptability of aggression.	.00

Individuals' demographic characteristics are used as individual covariates.

This includes sex, ethnicity, FARM status, baseline report card grades (quarter one in 2004), and a baseline composite of parent and teacher ratings of problem behaviors (fall 2004 administration). Ethnicity was dichotomized into two groups, African American, Hispanic, and American Indian students (0) and White and Asian American students (1). FARM status was coded as not receiving free or reduced

meals (0), receiving reduced meals (1), and receiving free meals (2). This was used as a proxy for socio-economic status.

Social competency rating data were skewed on the parent and teacher surveys, indicating low rates of problem behaviors and high levels of social skills. Data were inverted and normalized using a z-score transformation to reduce skewness. Parent and teacher scores were then averaged to create one composite of baseline problem behaviors. See Appendix D for correlations between this variable and the outcome measures.

The baseline report card grades were from the first quarter of the 2004-5 school year. This composite was the average of nine school report card items: Reading Level, Uses Reading Strategies, Comprehends what is Read, Writes Effectively, Listens and Speaks Effectively, Understands and Applies Mathematical Problem-Solving Processes, Understands and Applies Math Concepts, Understands and Applies Math Computational Processes, and Communicates Mathematical Concepts. Data were also skewed on these ratings. Scores were reflected, normalized using an inverse natural log transformation, and standardized by grade. See Appendix D for correlations between this variable and the outcome measures.

Treatment status was used as the sole level-2 predictor in benchmark analyses. In all nine spring 2006 analyses, the same individual and school predictors were used.

### *Analyses*

Because schools were assigned to treatment, students are nested within schools and cannot be considered independent participants. Therefore, data were analyzed using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) with

the individual at level-1 and the school at level-2. This ensures that the error will not be underestimated and that the statistical significance of effects will not be overestimated. Classrooms are not treated as a level because: (a) schools, not classrooms, were randomly assigned, (b) children do not retain the same classroom from year to year, (c) there is no significant difference in the amount of content delivered across grades, and (d) classrooms were not sampled (all were included) and so there is no sampling variance due to classroom sampling.

There are nine HLM benchmark equations in total, one for each of the included 2006 student survey scales. Following the advice of Raudenbush and Bryk (2002), all intercepts and independent (or predictor) variables were first tested for whether they have fixed or random effects on the spring score. Then all variables were added in grand-mean-centered form. Only those variables with significantly varying slopes were treated as random across schools. All covariates, whether or not they explain significant variance in the dependent variable, were retained in the equation. At level-2, intercept and slope coefficients in the level-1 equations were the dependent variables; treatment level was used as the sole level-2 predictor variable in benchmark analyses. The level-2 coefficient for treatment in the final estimation of fixed effects on the intercept in the level-1 equation indicates the size of the treatment effect. An example of an equation is as follows, using the Aggression final model:

Level One (Individual)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \bar{X}_{1..}) + \beta_{2j}(X_{2ij} - \bar{X}_{2..}) + \beta_{3j}(X_{3ij} - \bar{X}_{3..}) + \beta_{4j}(X_{4ij} - \bar{X}_{4..}) + \beta_{5j}(X_{5ij} - \bar{X}_{5..}) + r_{ij} \quad (1)$$

Level Two (School)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(W_{ij}) + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} \quad (3)$$

$$\beta_{2j} = \gamma_{20} \quad (4)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(W_{ij}) + u_{3j} \quad (5)$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}(W_{ij}) + u_{4j} \quad (6)$$

$$\beta_{5j} = \gamma_{50} + \gamma_{51}(W_{ij}) + u_{5j} \quad (7)$$

In Equation 1,  $Y_{ij}$  represents the spring 2006 student survey aggression scale,  $\beta_{1j}(X_{1ij} - \bar{X}_{1..})$  represents grand-mean-centered gender and its regression weight,  $\beta_{2j}(X_{2ij} - \bar{X}_{2..})$  represents grand-mean-centered ethnicity and its regression weight,  $\beta_{3j}(X_{3ij} - \bar{X}_{3..})$  represents grand-mean-centered FARM status and its regression weight,  $\beta_{4j}(X_{4ij} - \bar{X}_{4..})$  represents grand-mean-centered fall 2004 baseline composite of parent and teacher ratings of problem behaviors and its regression weight, and  $\beta_{5j}(X_{5ij} - \bar{X}_{5..})$  represents grand-mean-centered academic report card grades from quarter one of the 2004-5 school year and its regression weight.  $r_{ij}$  is an error term in the within-school model.  $u_{0j}$  is an error term in the level-2 model for the intercept in the level-1 equation,  $u_{3j}$ ,  $u_{4j}$ , and  $u_{5j}$  are the error terms for the randomly varying slopes,  $\gamma_{01}$  represents the treatment effect ( $W_{ij}$  is the treatment indicator) in Equation 2, and  $\gamma_{31}$ ,  $\gamma_{41}$ , and  $\gamma_{51}$  represent cross-level treatment interactions with student FARM status, baseline problem behaviors, and baseline grades respectively in Equations 5, 6, and 7.

*Sensitivity analyses.* Three a priori sensitivity analyses were performed to determine whether the findings are comparable to benchmark findings. These

analyses were used to provide alternatives to the benchmark analysis in the face of sample non-equivalence and attrition.

A threat to the validity of this study is sample non-equivalence despite randomization. White and Asian students were more likely to be assigned to treatment. This occurred because 75% of students in one control school are African American or Hispanic. Other schools had far fewer students in these ethnic/racial groups (30% of population in the matched school). The statistical test of equivalence was affected by this outlier school. To cope with this, a sensitivity analysis including only 10 schools was completed in which this school and its match were excluded from the analysis.

Another threat to validity was attrition, which occurred in this study in two ways. There was selective participation because not all parents gave consent for their children to participate in the survey and not all students responded to all questions on the survey. Lower rates of participation in the student survey were achieved at the individual level for children who received free or reduced meals (FARM) at school. When accounting for clustering at the school level however this lower rate of participation was not statistically significant. There was also no interaction between FARM status and treatment status when predicting whether a survey was received. In other words, FARM status does not have a different relationship with whether a student survey was received across treatment and control schools.

To address the concern that not all students completed certain scales on the survey (i.e., Aggression, Rebellious Behavior, Victimization, and Acceptability of Aggression), nine variables were created indicating whether a score was available for

each survey scale. These variables were used as the dependent variables in nine HLM analyses. Each individual predictor (grade level, FARM status, gender, ethnicity, baseline ratings of problem behaviors, and baseline report card grades) was added in group-mean-centered form and tested for significance and random variance. This was to determine if each predictor, analyzed separately, significantly predicted whether a survey scale score was available. For predictors with random variance, at level-2, slope coefficients in the level-1 equations were also dependent variables. This tested whether there was a treatment interaction with individual predictors in predicting missing data.

To cope with missing responses, sensitivity analyses using weights based on school, grade, ethnicity, and FARM status (Graham, Cumsille, & Elek-Fisk 2003) and imputation using an EM algorithm (Rubin, 1991) were used. Weighting and imputation re-constitutes the sample if people within groups were missing at random.

## Results

Based on the HLM analyses, no individual-level variable statistically significantly predicted whether a student received consent or submitted a student survey. Treatment status never predicted whether a survey was received or whether a scale score was available.

There were also no statistically significant relationships between the individual predictors and whether there was a scale score available for the Engagement in Learning, Empathy, Self-Restraint, Hostile Attribution Bias, and Altruism scales. The response rate on these scales was comparable or equal to the overall rate of survey submission.

Individual-level variables statistically significantly predicted whether a scale score was available on scales with lower response rates (Aggression, Rebellious Behavior, Victimization, and Acceptability of Aggression). Scale scores on Aggression, Rebellious Behavior, and Acceptability of Aggression were only available for 30-35% of the total third- and fifth-grade sample. Almost 50% of the total sample received scale scores for Victimization (refer to Table 2). This is far lower than the overall survey submission rate of 70%.

On the Aggression scale, response rates were predicted by the students' age/grade level, gender, and baseline report card grades and problem behaviors. Students' gender and baseline report card grades and problem behaviors were predictive of whether a score on the Rebellious Behavior scale was available. Baseline report card grades predicted whether students responded to the Victimization scale. Finally, gender and baseline ratings of problem behavior predicted whether students responded to the Acceptability of Aggression scale.

All individual variables were tested for random variance. Even when variables had randomly varying slopes, there were never significant interactions with treatment status. Therefore, the effects of individual predictors on missing data did not occur differentially across treatment and control schools.

Table 6 displays a summary of treatment effect estimates including the effect size and standard error for each dependent variable in the benchmark analysis and the three sensitivity analyses.



Table 6

*Second-Year Effect Sizes in Benchmark and Sensitivity Analyses*

Outcome	Benchmark Analyses		Analyses with 10 schools		Weighted Analyses		Imputed Analyses	
	Effect Size ( <i>ES</i> )	Standard Error ( <i>SE</i> )	<i>ES</i>	<i>SE</i>	<i>ES</i>	<i>SE</i>	<i>ES</i>	<i>SE</i>
Engagement in Learning	.11	.12	.06	.13	.10	.10	.09	.11
Empathy	.04	.15	.08	.16	.01	.13	.06	.13
Self-Restraint	.05	.13	.02	.14	.07	.11	.08	.10
Altruism	-.20	.12	-.27*	.08	-.22	.12	-.15	.11
Aggression	-.19	.13	-.21	.15	-.19	.12	-.10	.09
Rebellious Behavior	-.04	.16	-.13	.15	-.09	.10	-.08	.09
Acceptability of Aggression <sup>a</sup>	---	---	---	---	---	---	---	---
Hostile Attribution Bias	-.13	.11	-.11	.11	-.11	.10	-.13	.12
Victimization	-.16	.11	-.21*	.09	-.14	.11	-.16	.08

<sup>a</sup>Findings for Acceptability of Aggression are not reported because there was no

between-school variability.

\*  $p < .05$

Treatment effects for Acceptability of Aggression were not calculated because there is no between-school variability on this measure in the fully unconditional model ( $ICC = .00$ ). The main effect sizes for treatment on Empathy, Self-Restraint, and Rebellious Behavior are close to zero ( $ES = -.04$  to  $.05$ ). The effect sizes for treatment on Engagement in Learning, Hostile Attribution Bias, Aggression, and Victimization are not statistically significant but are in the desired direction; they are  $.11$  ( $SE = .12$ ,  $p = .39$ ),  $-.13$  ( $SE = .11$ ,  $p = .29$ ),  $-.19$  ( $SE = .13$ ,  $p = .19$ ), and  $-.16$  ( $SE = .11$ ,  $p = .21$ ) respectively. See Appendix E for detailed tables of the analyses on these

variables. The non-significant effect on Victimization was -.16 in the benchmark analysis, and ranged from -.14 to -.21 (nominal  $p < .05$ ) in the sensitivity analyses.

The effect size for treatment on Altruism is one-fifth of a standard deviation ( $ES = -.20$ ,  $SE = .12$ ,  $p = .13$ ), but not in the desired direction. Students in treatment schools report fewer helpful behaviors than students in control schools. At baseline, the coefficient was -.14. Pre-treatment differences also favored the control group (see Table 7).

The second research question for this study was whether Second Step is differentially effective based on free and reduced meal (FARM) status or baseline performance. This requires that the covariates (FARM, baseline social competence, or baseline academic report card grades) have randomly varying slopes. Treatment interactions were tested on all scales but Hostile Attribution Bias and Victimization because there were no randomly varying slopes for these measures. There were no significant treatment interactions with FARM or baseline measures of problem behaviors or academic grades on any of the remaining scales (See Appendix E for tables for each outcome).

Many of the effect sizes listed in Table 6 are noteworthy but some non-significant differences existed prior to the intervention. Because baseline scale scores could not be included as an individual covariate for the entire sample, control for baseline differences was lacking in the benchmark analyses. Table 7 displays the pre-treatment differences on each of the student survey scales.

Table 7

*Treatment Coefficients When Predicting Baseline Student Survey Scales*

Baseline Student Survey Scales	Effect Size	Standard Error	Degrees of freedom	Approx. <i>p</i> -value
Engagement in Learning	-.09	.07	10	.23
Empathy	.03	.08	10	.71
Self-Restraint	.04	.09	10	.66
Altruism	-.14	.09	10	.14
Hostile Attribution Bias	-.11	.15	10	.48
Aggression	-.22	.16	10	.19
Rebellious Behavior	-.03	.15	10	.87
Victimization	-.09	.07	10	.23
Acceptability of Aggression	.00	.12	10	.99

*Note.* Differences on the measures were not statistically significant. The magnitude of the effect sizes for Engagement and Altruism favor the control group. The magnitude of the effect sizes for Hostile Attribution Bias, Aggression, and Victimization favor the treatment group.

*Sensitivity analyses*

On the analysis which excluded the outlier school and its match, all main effect sizes were consistent with the benchmark analyses (See Table 6). Two outcomes reached statistical significance when only 10 schools were in the sample: there was a significant, negative effect on altruism (treatment students reported fewer helping behaviors) and a significant, negative effect on victimization (treatment students reported being victimized on fewer occasions).

To cope with missing data for students receiving free and reduced meals (FARM), sensitivity analyses using weights based on school, grade, ethnicity, and FARM status (Graham, Cumsille, & Elek-Fisk 2003) and imputation using an EM algorithm (Rubin, 1991) were used. Weighting and imputation re-constitutes the sample if people within groups were missing at random. Despite the issues with

attrition, effect sizes were consistent in size across benchmark and these two sensitivity analyses (See Table 6). The findings of the three sensitivity analyses support interpreting benchmark results.

To estimate the effects with increased control of the baseline differences, two analyses were completed. First, outcomes were assessed only for fifth-grade students who completed a baseline student survey. The baseline student survey scale scores were included as an individual covariate along with gender, FARM status, and ethnicity. Effect sizes for all scales were not statistically significant and many effect sizes appear smaller in magnitude than in the benchmark analyses. This implies that the effect sizes in the benchmark analyses may be in part due to pre-treatment differences.

Table 8

*Second-Year Effect Sizes When Third Graders Were and Were Not Included in the Analyses*

Outcome	Both Third and Fifth Grade		Fifth Grade Only	
	Effect Size ( <i>ES</i> )	Standard Error ( <i>SE</i> )	Effect Size ( <i>ES</i> )	Standard Error ( <i>SE</i> )
Engagement in Learning	.11	.12	.24	.17
Empathy	.04	.15	.12	.12
Self-Restraint	.05	.13	.09	.11
Altruism	-.20	.12	-.15	.09
Aggression	-.19	.13	-.05	.21
Rebellious Behavior	-.04	.16	-.03	.25
Acceptability of Aggression <sup>a</sup>	---	---	-.04	.17
Hostile Attribution Bias	-.13	.11	-.06	.15
Victimization	-.16	.11	-.07	.13

*Note.* None of the effect sizes reached statistical significance.

Second, an analysis with both third and fifth graders was completed including an aggregated baseline variable at level-2 as a predictor of the intercept (i.e. an aggregate of Engagement in Learning was added to the Engagement model and so on). It was expected that baseline aggregates would decrease standard errors and increase power (Bloom, Richburg-Hayes, & Black, 2005). When this analysis was performed, still no significant findings emerged and most treatment effect sizes decreased. Like the analysis with just fifth graders, baseline differences on the measures were adjusted with this technique. These supplementary analyses provide evidence that Second Step did not have effects in this sample after two years of implementation.

## Discussion

Second Step is an easy-to-use, manualized curriculum that integrates many features found in other prevention programs that have been found to be efficacious. There are lesson kits for pre-K, elementary, and middle school teachers. Social skills are explicitly taught to students, including empathy, anger management, impulse control, and problem-solving skills (Committee for Children, n.d.). The use of a cognitive-behavioral approach that integrates modeling, teaching, and feedback about these skills, is expected to help children to think differently about social situations so that behavior will change (Frey et al., 2000; Hahn et al., 2007).

Second Step is a widely used program despite a lack of empirical support for its effectiveness. It is marketed to all schools as a research-based program that enhances school climate, decreases discipline concerns, and provides students with increased confidence and responsibility for their behavior (Committee for Children,

n.d.). These claims for success are based on the same studies that were reviewed in the present study.

To date, this is only the second experimental trial evaluating Second Step. An earlier study by Grossman et al. (1997) found some effects on observations, but these effects were not confirmed by parent or teacher survey data. Effect sizes were not calculated for any outcomes in the Grossman et al. study. Information to calculate the effect sizes based on observations was not provided. Information to calculate effect sizes based on parent and teacher ratings was available; the largest effects were -.05 on ratings of aggression and delinquency. The remaining scales had effects of .00. All effects were consistent across parent and teacher ratings and were non-significant.

After one year of the present experiment evaluating Second Step, Harak (2006) reported significant effects ( $ES = -.15, p = .05$ ) on only one of 11 student self-report scales. Students in the treatment schools reported being more attentive and engaged in classroom instruction after one year of implementation than control students.

Frey et al. (2005) asserted that random assignment was used in an additional study, but random assignment was not carried out in at least four study schools. Main effect sizes on teacher ratings of social behavior were not provided. The vague descriptions of the design and outcomes make it difficult to conclude that Frey et al. provides true experimental evidence of the effectiveness of Second Step. Effects found on ratings in the first year were not replicated in the second year.

After two years of implementation of Second Step in six suburban schools, there are no statistically significant results on student self-reports of engagement in

learning, prosocial skills, or problem behaviors. As in the Frey et al. (2005) study, first year findings of significance were not replicated. While effect sizes in the present study appear noteworthy, these effects mirror the non-significant differences present at baseline. This was confirmed by the two sensitivity analyses which controlled for baseline performance. The benchmark model findings of non-significance were either replicated or effects were decreased in magnitude across all sensitivity analyses; providing no evidence that Second Step affects the outcomes examined.

Despite the lack of evidence, the Committee for Children (n.d.) indicates that their four programs are used in 21 countries, 25,000 schools, and are delivered to 7 million children. This highlights the importance of rigorous evaluations of prevention and social skills programs generally and Second Step specifically. Of the experimental evaluations of Second Step, this is the first completed by independent researchers.

### *Strengths*

Among the strengths of this research is the random assignment of schools to treatment and control; a universal, high quality implementation of the Second Step curriculum for two years; and a large sample. Grossman et al. (1997) used a smaller sample, did not take a universal approach, and implemented for only one year. Even with these improvements, effects in this evaluation are similar to earlier findings by Harak (2006), Grossman et al., and Frey et al. (2005).

Based on teacher self-reports, all schools implemented over 80% of the total lessons for their grade-level. Within the lessons, all but one school implemented 80% of the key elements. Students were instructed by their classroom teachers about one

time per week. The implementation took place across two years. As compared to national norms, this is a strong implementation (D.C. Gottfredson & G.D. Gottfredson, 2002). On average, direct services programs are implemented for less than a semester, are delivered to about one-third of students, and programs are delivered by some (not all) teachers and school support staff (D.C. Gottfredson & G.D. Gottfredson, 2002).

Qualitatively, school counselors in the treatment schools were enthusiastic about Second Step. Despite the presentation of pervasively non-significant results the school counselors still felt strongly that Second Step was effective (Gottfredson et al., 2008). Specifically, counselors felt that students had a better understanding of others and how to resolve conflicts. After the evaluation of Second Step was completed, the intervention schools continued and control schools began implementing Second Step. Schick and Cierpka (2005) encountered similar attitudes from the educators who took part in their study.

### *Limitations*

Despite a rigorous design, some weakness should be noted. First, the outcomes of this study are reliant on a student self-report measure. Student self-reports usually detect small effect sizes (Wilson & Lipsey, 2007). It is possible that elementary-aged students, particularly third graders, were too young to complete such a survey. Based on analyses of item non-response, age/grade level did not statistically significantly predict overall survey submission or response rate on 8 out of 9 scales. The significantly different response rates on the aggression scale may have been to sensitivity of the scale. While falling late in the survey, it was followed by three



scales of items which did not have differential response rates based on grade. This suggests that third graders were able to respond to this survey at the same rate as the fifth grade students overall.

Because only 12 schools were included in the study, power to detect effects is relatively low. Based on meta-analyses, it was expected that an intervention like Second Step would yield effect sizes of .20. When effects were examined for fifth graders only, the treatment effect on Engagement in Learning was .24, but was not statistically significant. It is possible that with more schools, this effect would have reached significance. On the other hand, almost no other effect sizes were of this magnitude. When baseline was properly controlled, none of the effect sizes but Engagement were equal to or larger than .20. The next largest effect ( $ES=-.15$ ) was on the Altruism scale, but indicated a non-desired effect in which treatment students reported being less helpful than control students.

Next, the sample used was not necessarily a sample in need of this universal intervention. The baseline rates of problem behavior were low across multiple measures. Data from parents, teachers, students, and report cards were skewed to reflect positive behaviors and achievement. Almost every student survey scale needed to be transformed to normalize the distribution of scores. This leaves open the possibility of ceiling effects. To examine this possibility, interaction effects with the individual predictors were tested but not detected on any of self-report scales. In other words, the treatment did not differentially affect the students who were initially rated as displaying higher rates of problem behaviors or lower achievement. This implies that ceiling effects are not the sole reason why treatment effects were not detected.

The sample also is not representative of many school systems in the United States. Of the twelve schools involved in this study, eight have a small ethnic minority population. The two schools with the largest population of ethnic minorities were not an even match (one had 75% students minority and the other had 33%). The inclusion of these two schools in the study created sample non-equivalence. When the outcomes were evaluated without these two schools however the non-equivalence disappeared but the findings remained the same.

A final issue is attrition because not all students responded to all of the survey items. In particular, individual demographics and baseline data predicted whether a student responded to the Aggression, Rebellious Behavior, Victimization, and Acceptability of Aggression scales. While this resulted in low response rates (30-50% of the entire third and fifth grade sample), this response pattern occurred consistently in both treatment and control schools. This is important because one may hypothesize that the lack of statistically significant effects on these student self-report scales were due to sample non-equivalence. It appears that students' non-response rates on certain scales related to risk factors, but this occurred equally in both the treatment and control schools. This finding rules out the possibility that non-significant findings were the result of more at-risk students completing survey scales in the treatment than control schools.

#### *Implications for Future Research and Practice*

When considering these results in the context of earlier findings by Harak (2006), Grossman et al. (1997), and Frey et al. (2005), it does not appear that Second Step is effective in decreasing student aggression or increasing social skills. There are

weaknesses of each study however that could be considered in future research on Second Step. First, a concern for the present study is whether the students who did not receive consent (roughly 30% of the sample) were the ones who needed the intervention the most. Based on analyses of demographic information (male students, those from a lower socio-economic status, and ethnic minorities), higher risk students were not statistically less likely to receive consent. It could be the case though that the students who were not given consent were the students who are lower school achievers or whose parents and teachers would have rated them poorly. Without consent, the data to explore this hypothesis are not available.

In addition, the low response rates on the four scales pertaining to aggression and rebellious behavior are of concern. With only 30% of the sample providing information, it is difficult to say what the impacts of Second Step truly were on negative behaviors. A positive finding is that differential response based on demographic characteristics did not vary across treatment and control schools. One way to improve this in future studies would be for the researchers to be more involved in the administration of the student surveys.

A second area of improvement would be implementation monitoring. Observations could be used to verify the validity of the contemporaneous implementation logs. McKenna (2005) found that in teacher self-reports of implementation of the Instructional Consultation were discrepant from observations of consultation sessions. It is possible that the teacher self-reports were accurate in the quantity, but not the quality, of implementation and that observations would provide

more insight into this hypothesis. This would however be a costly and time-consuming endeavor.

Also relating to implementation is the possibility that not all students received the intervention as expected. Although teachers presented lessons to their classes, there are no data about the number of students present for the lessons. Despite very high student attendance rates at these schools, it is possible that because of the organization and scheduling in schools, lessons may have overlapped with pullout activities such as special education, Title I, or English for Speakers of Other Languages (ESOL) services. Title I and ESOL services would be prevalent in the potentially highest need schools which serve students who are culturally, linguistically, and ethnically diverse and come from homes of low socio-economic status.

It is likely that Second Step is not a necessary component of the social and behavioral curriculum at schools that are generally well-functioning and serve an affluent student population. Despite differences in the formality, materials, and activities of the treatment and control schools in the present study, the control schools still had something in place. Research shows that all schools have a large number of programs in place at any given time; with a median of 14 prevention program (D.C. Gottfredson & G.D. Gottfredson, 2001). It is possible that when compared to a “control” such as this, Second Step is not effective.

One may wonder whether Second Step would be more effective in more diverse, higher need schools. The findings of this study cannot speak directly to that. Perhaps schools that lack beneficial organizational structure would need a program

like Second Step. But one must wonder whether it would be implemented well enough to be effective. This raises the question, “Can prevention work where it is needed most?” (D.C. Gottfredson, G.D. Gottfredson, & Skroban, 1998). As D.C. Gottfredson et al. (1998) concluded, whether prevention can work all depends on a school’s capacity and willingness to fully adopt programs of change.

## Author Note

This research is supported in part by grant number R305L030002A from the Institute of Education Sciences, U.S. Department of Education. Opinions expressed do not necessarily reflect the opinions or policies of the U.S. Department of Education. The development of the data used in this research would not have been possible without the members of the Character Education Inquiry research team at the University of Maryland (Gary D. Gottfredson, Amy Silverman, Allison J. Nebbergall, and Kyra Richardson) or the Anne Arundel County School System Administration (Lucia Martin and Debbie Wooleyhand) and Counselors in each of the twelve research schools (Sharon Atkinson, Kristin Clute, Cindi Cramer, Rachel Dressin, Beth Edelstein, Jenn Elsis, Holly Kleiderlein, Patricia Lynch, Joyce O'Connell, Carole Peacock, Connie Poussard, Aileen Selznick, Dana Smith, Frances Walker, and Teresa Wismer).

Correspondence concerning this manuscript should be addressed to Elise Harak, Counseling and Personnel Services, University of Maryland – College Park, College Park, MD 20742. Email: [eharak@umd.edu](mailto:eharak@umd.edu).

## Footnotes

<sup>1</sup> The SACD program includes a national study conducted by IES contractor Mathematica Policy Research (MPR, John Burghart & Peter Schochet); complementary studies conducted by Larry Aber (NYU), Leonard Bickman (Vanderbilt), Mark Fraser (UNC), Brian Flay (UIC), Gary Gottfredson (UMD), Debbie Johnson (Children's Institute), and William Pelham (SUNY Buffalo); and with participation by the Centers for Disease Control and Prevention (CDC).

Appendix A  
Example of an Implementation Log

**Grade 1 Second Step Implementation Record**  
**Unit I: Empathy Training**

Date Month	Lesson Delivered		Lesson 1: Empathy Training-Skill Overview		In conducting this lesson, did you . . .
	Day	Day	Yes	No	
Sep	0	0	<input type="checkbox"/>	<input type="checkbox"/>	1. Distribute take-home letter 1: Introduction to <i>Second Step</i> ?
Oct	1	1	<input type="checkbox"/>	<input type="checkbox"/>	2. Use the photo to introduce what will be learned?
Nov	2	2	<input type="checkbox"/>	<input type="checkbox"/>	3. Have students generate rules for during the lessons?
Dec	3	3	<input type="checkbox"/>	<input type="checkbox"/>	4. Have students recite the rules?
Jan	4	4	<input type="checkbox"/>	<input type="checkbox"/>	5. Have a discussion about the weekend to practice using rules?
Feb	5	5	<input type="checkbox"/>	<input type="checkbox"/>	6. Praise students who followed the rules?
Mar	6	6	<input type="checkbox"/>	<input type="checkbox"/>	7. Discuss times when students can use the rules?
Apr	7	7	<input type="checkbox"/>	<input type="checkbox"/>	8. Do the listening game?
May	8	8	<input type="checkbox"/>	<input type="checkbox"/>	9. List feeling names and a face to match on poster paper?
Jun	9	9	<input type="checkbox"/>	<input type="checkbox"/>	10. Hang the poster with the rules on the wall?

Comments about lesson 1:

*Note.* Format shown here is not an exact replica of formatting used. Text is the same.



## Appendix B

### First-Year Implementation Findings (Taken from Harak, 2006)

An ANOVA revealed that there were significant differences between schools on the percentage of logs submitted ( $F = 6.31, p = 0.00$ ) and the percentage of curriculum delivered ( $F = 5.50, p = 0.00$ ) but not between grades. Of the six schools, five of the six schools reached “successful implementation” (80% or more of lessons delivered) as determined by implementation standards created by the research team in collaboration with school personnel. Four of the six schools did not implement 80% of the total curriculum however. Implementation was measured as a dichotomous variable; yes or no for whether the log and each element of the lesson were implemented.

## Appendix C

### Scale Descriptions (Taken from Harak, 2006)

*Engagement in Learning.* This scale was adapted by IES from Furrer and Skinner (2003). Composed of nine questions such as “I try to do well in school” and “I pay attention in class,” the Likert-type scale for each item ranged from “agree a lot” to “disagree a lot.” The alpha reliability for this scale in the fall administration is 0.75; the fall 2004 to spring 2005 stability correlation is 0.47. [The fall 2004 to spring 2006 stability correlation is .34]. This is the most distal potential outcome for *Second Step*, but if the program is successful at decreasing aggression and rebellion, the students may become less distracted by the environment and more engaged in school and learning.

*Empathy.* This scale was adapted by IES from Funk, Elliott, Bechtoldt, Pasold, and Tsavoussis (2003). Composed of ten questions such as “I understand how other kids feel” and “Other people's problems really bother me,” the Likert-type scale for each item ranged from “YES!” to “NO!” The alpha reliability for this scale is 0.79; the [first-year] stability coefficient is 0.51. [The fall 2004 to spring 2006 stability correlation is .37]. Perspective-taking is directly taught to the students in the beginning of most lessons; students are shown a picture and read a vignette and asked how the child in the scenario is feeling and how the students know.

*Self-Restraint.* This scale was constructed for the purpose of the present research. Composed of seven questions such as “If two kids are fighting, someone should stop it” and “I know a way to calm down when I start to get angry,” the

Likert-type scale for each item ranged from “YES!” to “NO!” The alpha reliability for this scale is 0.84; the [first-year] stability coefficient is 0.50. [The fall 2004 to spring 2006 stability correlation is .38]. The skills of restraint are directly taught in the anger management and problem-solving units of *Second Step*.

*Hostile Attribution Bias.* Some questions relating to hostile attribution bias were included in an IES Victimization Scale. The Hostile Attribution Bias Scale was constructed using these IES questions and additional questions constructed for the purpose of the present research. Composed of six questions such as “If some kids get candy and I don't get any, I was probably left out on purpose” and “When kids hurt my feelings, they do it to be mean,” the Likert-type scale for each item ranged from “YES!” to “NO!” The alpha reliability for this scale is 0.73; the [first-year] stability coefficient is 0.39. [The fall 2004 to spring 2006 stability correlation is .32]. A skill set taught in each *Second Step* lesson is the assessment of a situation; with this skill students should not attribute ambiguous situations to hostile intentions.

*Altruism.* This scale was adapted by IES from Solomon, Battistich, Watson, Schaps, and Lewis (2000). Composed of eight questions such as “At school or someplace else, I helped someone who was being picked on” and “At school or someplace else, I got help for someone who was hurt,” the Likert-type scale for each item included “never,” “once or twice,” “a few times,” and “many times.” The alpha reliability for this scale is 0.86; the [first-year] stability coefficient is 0.47. [The fall 2004 to spring 2006 stability correlation is .40]. *Second Step* addresses doing helpful and nice things for others as lessons.

*Aggression.* This scale was adapted by IES from Orpinas and Frankowski (2001). Composed of six questions such as “I left out another kid on purpose,” “I said that I would hit a kid at school,” and “I pushed, shoved, or hit a kid from school,” the Likert-type scale for each item included “never,” “once or twice,” “a few times,” and “many times.” The alpha reliability for this scale is 0.88; the [first-year] stability coefficient is 0.40. [The fall 2004 to spring 2006 stability correlation is .33]. Aggression is specifically targeted by the *Second Step* curriculum in the teaching of alternative, non-angry and aggressive, solutions to problems.

*Frequency of Rebellious Behavior.* This scale was adapted by IES from Loeber and Dishion (1983). Composed of six questions such as “I took something from someone at school that did not belong to me” and “I copied other students' homework or copied off of other students' tests,” the Likert-type scale for each item included “never,” “once or twice,” “a few times,” and “many times.” The alpha reliability for this scale is 0.83; the [first-year] stability coefficient is 0.27. [The fall 2004 to spring 2006 stability correlation is .18]. Certain lessons in the curriculum specifically address issues of cheating, lying, and stealing; students who receive the program should show decreased endorsement of engaging in these acts.

*Victimization.* This scale was adapted by IES from Orpinas and Kelder (1995). Composed of six questions such as “A kid from school pushed, shoved, or hit me” and “A kid from school called me a bad name,” the Likert-type scale for each item included “never,” “once or twice,” “a few times,” and “many times.” Some questions were moved from the IES victimization to the Hostile Attribution Scale, so the Victimization scale is shorter than the scale used in the national study. The alpha

reliability for this scale is 0.87; the [first-year] stability coefficient is 0.52. [The fall 2004 to spring 2006 stability correlation is .30]. Because *Second Step* is intended to be a violence prevention program, if it is effective and aggression decreases, students should endorse that they experience less victimization in those schools.

*Acceptability of Aggression.* This scale was adapted by IES from Huesmann and Guerra (1997). Composed of eight questions such as “It is OK to yell at others and say bad things” and “It is OK to take your anger out on others by using physical force,” the Likert-type scale for each item included “really wrong,” “sort of wrong,” “sort of OK,” and “perfectly OK.” The alpha reliability for this scale is 0.88; the [first-year] stability coefficient is 0.30. [The fall 2004 to spring 2006 stability correlation is .28]. It is expected that since alternative, non-aggressive, solutions are taught in *Second Step*, those students who received the treatment will decrease their endorsement about how acceptable aggression is as a solution to a problem.

## Appendix D

### Correlations Between Baseline Student Survey Scales and Spring 2006 Outcomes

Baseline Student Survey Scales	Baseline average academics score	Composite of parent and teacher rating
Engagement in Learning	-0.24*	0.32*
Empathy	-0.09*	0.18*
Self-restraint	-0.07	0.24*
Attribution Bias	-0.20*	0.22*
Altruism	-0.08	-0.02
Aggression	-0.17*	0.33*
Rebellious Behavior	-0.26*	0.29*
Victimization	-0.20*	0.30*
Acceptability of Aggression	-0.09	0.07

*Note.* Baseline average academics score is the standardized academic grades from the 2004-5, quarter one report cards. Composite of parent and teacher rating is the average of the fall 2004 ratings completed by parents and teachers about student problem behaviors.

\*  $p < .05$ , \*\*  $p < .01$

Appendix E  
HLM Outcome Tables

Table E1

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Engagement in Learning Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	.01	.09	10	.92
Treatment status (0 = control, 1 = treatment)	.11	.12	10	.39
Individual sex (0 = female, 1 = male)	-.14	.15	10	.38
Sex x treatment interaction	.15	.21	10	.49
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.04	.14	802	.79
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.04	.09	802	.68
Quarter 1 (2004) academic report card grades	.05	.04	802	.17
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	-.38	.08	10	.00*
Baseline problem behaviors x treatment interaction	.14	.11	10	.24
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.84	--	--	--
Level-2 intercept	.03	10	29.64	.00
Sex slope	.07	10	23.58	.01
Baseline problem behaviors slope	.02	10	20.53	.02
<u>Proportion of variance explained</u>				
Proportion of tau explained	.00			
Proportion of sigma-squared explained	.13			

*Note.* SE = standard error.

<sup>a</sup>A positive coefficient indicates an increase in engagement in learning.

\*  $p < .05$

Table E2

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Empathy Made**by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	-.04	.11	10	.69
Treatment status (0 = control, 1 = treatment)	.04	.15	10	.81
Individual sex (0 = female, 1 = male)	.22	.07	801	.00*
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.22	.15	801	.14
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	-.10	.18	10	.59
FARM x treatment interaction	.12	.23	10	.60
Quarter 1 (2004) academic report card grades	-.02	.04	801	.63
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	.17	.04	801	.00*
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.93	--	--	--
Level-2 intercept	.05	9	35.06	.00
FARM slope	.04	9	14.92	.09
<u>Proportion of variance explained</u>				
Proportion of tau explained	.00			
Proportion of sigma-squared explained	.04			

*Note.* SE = standard error.

<sup>a</sup>A positive coefficient indicates an increase in empathy.

\*  $p < .05$



Table E3

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Self-Restraint**Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	.05	.09	10	.59
Treatment status (0 = control, 1 = treatment)	.05	.13	10	.70
Individual sex (0 = female, 1 = male)	-.12	.07	789	.08
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	.24	.14	789	.09
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.37	.17	10	.06
FARM x treatment interaction	-.32	.22	10	.18
Quarter 1 (2004) academic report card grades	.04	.04	789	.33
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	-.20	.07	10	.02*
Baseline problem behaviors x treatment interaction	.04	.10	10	.68
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.85	--	--	--
Level-2 intercept	.03	9	30.46	.00
FARM slope	.05	9	11.67	.23
Baseline problem behaviors slope	.01	9	17.56	.04
<u>Proportion of variance explained</u>				
Proportion of tau explained	.00			
Proportion of sigma-squared explained	.14			

*Note.* SE = standard error.<sup>a</sup>A positive coefficient indicates an increase in self-restraint.\*  $p < .05$

Table E4

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Altruism Made**by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	.02	.06	10	.78
Treatment status (0 = control, 1 = treatment)	-.20	.12	10	.13
Individual sex (0 = female, 1 = male)	-.09	.07	799	.21
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.08	.14	799	.58
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.18	.08	799	.05*
Quarter 1 (2004) academic report card grades	-.10	.04	799	.01*
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	-.05	.05	10	.35
Baseline problem behaviors x treatment interaction	-.14	.09	10	.16
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.90	--	--	--
Level-2 intercept	.03	10	33.12	.00
Baseline problem behaviors slope	.01	10	17.11	.07
<u>Proportion of variance explained</u>				
Proportion of tau explained	.09			
Proportion of sigma-squared explained	.08			

*Note.* SE = standard error.<sup>a</sup>A negative coefficient indicates a decrease in helping behaviors.\*  $p < .05$

Table E5

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Attribution Bias**Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	.02	.08	10	.84
Treatment status (0 = control, 1 = treatment)	-.13	.11	10	.29
Individual sex (0 = female, 1 = male)	.05	.07	792	.46
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.13	.14	792	.38
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.01	.08	792	.94
Quarter 1 (2004) academic report card grades	-.11	.04	792	.01*
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	.15	.04	792	.00*
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.87	--	--	--
Level-2 intercept	.02	10	27.01	.00
<u>Proportion of variance explained</u>				
Proportion of tau explained	.58			
Proportion of sigma-squared explained	.09			

*Note.* SE = standard error.

<sup>a</sup>A negative coefficient indicates a decrease in interpreting hostile intentions in ambiguous social situations.

\*  $p < .05$

Table E6

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Aggression**Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	-.02	.10	10	.87
Treatment status (0 = control, 1 = treatment)	-.19	.13	10	.19
Individual sex (0 = female, 1 = male)	.09	.10	378	.35
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.39	.17	378	.02*
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.21	.23	10	.39
FARM x treatment interaction	.14	.29	10	.64
Quarter 1 (2004) academic report card grades	.09	.09	10	.32
Baseline grades x treatment interaction	-.06	.12	10	.62
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	.22	.11	10	.07
Baseline problem behaviors x treatment interaction	-.03	.15	10	.82
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.78	--	--	--
Level-2 intercept	.02	6	11.25	.08
FARM slope	.10	6	13.76	.03
Baseline grades slope	.01	6	7.70	.26
Baseline problem behaviors slope	.03	6	16.22	.01
<u>Proportion of variance explained</u>				
Proportion of tau explained	.67			
Proportion of sigma-squared explained	.17			

*Note.* SE = standard error.

<sup>a</sup>A negative coefficient indicates a decrease in verbal, physical, and relational aggression.

\*  $p < .05$

Table E7

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Rebellious**Behavior Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	-.12	.12	10	.32
Treatment status (0 = control, 1 = treatment)	-.04	.16	10	.80
Individual sex (0 = female, 1 = male)	.15	.11	305	.15
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.06	.32	10	.86
Ethnicity x treatment interaction	.21	.50	10	.68
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.17	.35	10	.62
FARM x treatment interaction	-.01	.44	10	.99
Quarter 1 (2004) academic report card grades	.00	.11	10	.98
Baseline grades x treatment interaction	-.11	.15	10	.48
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	.12	.06	305	.04*
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.79	--	--	--
Level-2 intercept	.04	3	14.46	.00
Ethnicity slope	.23	3	6.86	.08
FARM slope	.17	3	10.43	.02
Baseline grades slope	.02	3	10.33	.02
<u>Proportion of variance explained</u>				
Proportion of tau explained	.36			
Proportion of sigma-squared explained	.16			

Note. SE = standard error.

<sup>a</sup>A negative coefficient indicates a decrease in rebellious behavior such as stealing, cheating, and skipping school.

\*  $p < .05$

Table E8

*Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Victimization**Made by Students, Third and Fifth Grade Sample*

<u>Fixed effects</u>				
Variable	Coefficient <sup>a</sup>	SE	df	p
Intercept	.01	.08	10	.88
Treatment status (0 = control, 1 = treatment)	-.16	.11	10	.21
Individual sex (0 = female, 1 = male)	.10	.08	549	.23
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.07	.17	549	.66
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.09	.09	549	.33
Quarter 1 (2004) academic report card grades	-.03	.05	549	.49
Individual fall 2004 composite of parent and teacher ratings of problem behaviors	.24	.05	549	.00*
<u>Random effects</u>				
Variance component	Variance	df	X <sup>2</sup>	p
Level-1 error	.91	--	--	--
Level-2 intercept	.02	10	18.02	.05
<u>Proportion of variance explained</u>				
Proportion of tau explained	.36			
Proportion of sigma-squared explained	.07			

Note. SE = standard error.

<sup>a</sup>A negative coefficient indicates a decrease in the frequency that students were victims of verbal, physical, and relational aggression.

\*  $p < .05$

## References

- Battistich, V., Solomon, D., Watson, M., Solomon, J., & Schaps, E. (1989). Effects of an elementary program to enhance prosocial behavior on children's cognitive-social problem-solving skills and strategies. *Journal of Applied Developmental Psychology, 10*, 147-169.
- Beelman, A., Pfingsten, U., & Losel, F. (1994). Effects of training social competence in children: A meta-analysis of recent evaluation studies. *Journal of Clinical Child Psychology, 23*, 260-271.
- Bloom, H.S., Richburg-Hayes, L., & Black, A.R. (2005). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions*. MDRC Working Papers on Research Methodology.
- Botvin, G.J. & Kantor, L.W. (2000). Preventing alcohol and tobacco use through life skills training. *Alcohol Research and Health, 24*, 250-257.
- Botzer, E.A. (2003). An evaluation of the effectiveness of the *Second Step* violence prevention curriculum for third-grade students. *Dissertation Abstracts International, 64* (4), 1171A.
- Committee for Children (1991). *Second Step: A Violence Prevention Curriculum, first-fifth grade*. Seattle, WA: Committee for Children.
- Committee for Children (n.d.). Retrieved June 18, 2008 from [www.cgchildren.org](http://www.cgchildren.org).
- Conduct Problems Prevention Research Group (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648-657.

- Cooke, M.B, Ford, J., Levine, J., Bourke, C., Newell, L., & Lapidus, G. (2007). The effects of city-wide implementation of Second Step on elementary school student's prosocial and aggressive behaviors. *The Journal of Primary Prevention, 28*, 93-115.
- Durlak, J.A. & Wells, A.M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25*, 115-152.
- Elias, M. J., Gara, M. A., Ubriaco, M., Rothbaum, P.A., Clabby, J.F., & Schuyler, T. (1986). Impact of a preventive social problem solving intervention on children's coping with middle school stressors. *American Journal of Community Psychology, 14*, 259-275.
- Elias, M. J., Gara, M. A., Schuyler, T. F., & Branden-Muller, L. R. (1991). The promotion of social competence: Longitudinal study of a preventive school-based program. *American Journal of Orthopsychiatry, 61*, 409-417.
- Flannery, D.J, Liau, A.K, Powell, K.E., Vazsonyi, A.T., Guo, S., Atha, H., et al. (2003). Initial behavior outcomes for the Peace Builders universal school-based violence prevention program. *Developmental Psychology, 39*, 292-308.
- Flay, B. R., & Allred, C. G. (2003). Long-term effects of the Positive Action program. *American Journal of Healthy Behavior, 27(1)*, 6–21.
- Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). *Second Step: Preventing aggression by promoting social competence. Journal of Emotional and Behavioral Disorders, 8*, 102-112.



- Frey, K. S., Nolen, S.B., Edstrom, L.V.S., & Hirschstein, M. K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Applied Developmental Psychology, 26*, 171-200.
- Funk, J., Elliott, R., Bechtoldt, H., Pasold, T. & Tsavoussis, A. (2003). The Attitudes Toward Violence Scale: Child Version. *Journal of Interpersonal Violence, 18*, 186-196.
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95*, 148-162.
- Gottfredson, D.C. & Gottfredson, G.D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency, 39(1)*, 3-35.
- Gottfredson, D.C., Gottfredson, G.D, & Skroban, S. (1998). Can prevention work where it is needed most? *Evaluation Review, 22*, 315-340.
- Gottfredson, G. D. (1987). American education—American delinquency. *Today's Delinquent, 6*, 5-70.
- Gottfredson, G. D. (1999). *The Effective School Battery: User's manual*. Ellicott City, MD: Gottfredson Associates.
- Gottfredson, G. D., & Gottfredson, D. C. (1999). *Development and applications of theoretical measures for evaluating drug and delinquency prevention programs: Technical manual for research editions of What About You (WAY)*. Ellicott City, MD: Gottfredson Associates.

- Gottfredson, G. D., & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Educational and Psychological Consultation, 12*, 313-344.
- Gottfredson, G. D., Gottfredson, D. C., Czeh, E. R., Cantor, D., Crosse, S., & Hantman, I. (2000). National study of delinquency prevention in schools. Ellicott City, MD: Gottfredson Associates. (ERIC No. ED 459 409)
- Gottfredson, G., Harak, E., Nebbergall, A., Nese, J., & Shaw, F. (2008, June). An experimental evaluation of an elementary school social competency instruction program. Symposium at the Social and Character Development Research Program Consortium Meeting, Arlington, VA.
- Gottfredson, G. D., Jones, E. M., Gore, T. W. (2002). Implementation and evaluation of a cognitive-behavioral intervention to prevent problem behavior in a disorganized school. *Prevention Science, 3*, 43-56.
- Graham, J.W.; Cumsille, P.E.; & Elek-Fisk, E. (2003). Methods for handling missing data. In J.A. Schinka, W.F. Velicer, I.B. Weiner (eds.), *Handbook of Psychology: Volume 2 Research Methods in Psychology* (pp. 87-114). Hoboken, New Jersey: John Wiley & Sons Inc.
- Grossman, D.C., Neckerman, H.J., Koepsell, T.D., Liu, P.Y., Asher, K.N., Beland, K., et al. (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.

- Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., et al. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine, 33*(2S), 114-129.
- Hall, B. W., & Bacon, T. P. (2005). Building a foundation against violence: Impact of a school-based prevention program on elementary students. *Journal of School Violence, 4*(4), 63–83.
- Harak, E.T. (2006). *First-year self-report outcomes of a character education experiment with elementary students*. Unpublished masters thesis, University of Maryland, College Park.
- Harris, P.A. (1998). Teaching conflict resolution skills to children: A comparison between a curriculum based and a modified peer mediation program. *Dissertation Abstracts International, 59* (9), 3397B.
- Hawkins, J.D., Catalano, R.F., Kosterman, R., Abbott, R. & Hill, K.G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics and Adolescent Medicine, 153*, 226-234.
- Hawkins, J.D., Catalano, R.F., & Miller, J.Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin, 112*, 64-105.
- Hirschi, T. (1969). *Causes of delinquency*. Berkeley, CA: University of California Press.

- 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.
- Ialongo, N.S., Werthamer, L., Kellam, S.G., Brown, C.H., Wnag, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*, 599-641.
- Institute of Education Sciences (n.d.). Retrieved April 3, 2007 from <http://www.whatworks.ed.gov/Topic.asp?tid=12&ReturnPage=default.asp>
- Institute of Education Sciences (n.d.). Retrieved May 18, 2007 from [http://www.whatworks.ed.gov/PDF/Intervention/techappendix12\\_250.pdf](http://www.whatworks.ed.gov/PDF/Intervention/techappendix12_250.pdf)
- Institute of Education Sciences (n.d.). Retrieved May 18, 2007 from [http://www.whatworks.ed.gov/PDF/Intervention/techappendix12\\_208.pdf](http://www.whatworks.ed.gov/PDF/Intervention/techappendix12_208.pdf)
- Leming, J. S. (2000). Tell me a story: An evaluation of a literature-based character education programme. *Journal of Moral Education, 29*, 413-427.
- Lillenstein, J.A. (2001). *Efficacy of a social skills training curriculum with early elementary students in four parochial schools*. Unpublished doctoral dissertation, Indiana University of Pennsylvania.
- Linares, L.O, Rosbruch, N, Stern, M.B., Edwards, M.E., Walker, G., Abikoff, H.B., et al. (2005). Developing Cognitive-Social-Emotional competencies to enhance academic learning. *Psychology in the Schools, 42*, 405- 417.
- Loeber, R., & Dishion, T.J. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin, 94*, 325-382.

- McDonald, C.A. (2001). *The relationship between Second Step curriculum and student office referrals, attendance, and academic achievement*. Unpublished doctoral dissertation, Saint Louis University.
- McMahon, S.D., Washburn, J. Felix, E.D., Yakin, J., & Childrey, G. (2000). Violence prevention: Program effects on urban preschool and kindergarten children. *Applied & Preventive Psychology, 9*, 271-281.
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701.
- Nebbergall, A.J. (2007). *Assessment of social competence and problem behavior: The psychometric properties of a social competency rating scale*. Unpublished masters thesis, University of Maryland, College Park.
- Newcomb, M.D., Abbott, R.D., Catalano, R.F., Hawkins, J.D., Battin-Pearson, S., & Hill, K. (2002). Mediation and Deviance Theories of Late High School Failure: Process Roles of Structural Strains, Academic Competence, and General versus Specific Problem Behaviors. *Journal of Counseling Psychology, 49*, 172-186.
- Olweus, D. (1997). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education, 12*, 495-510.
- Orpinas, P., & Frankowski, R. (2001). The Aggression Scale: A self-report measure of aggressive behavior for young adolescents. *Journal of Early Adolescence, 21*, 50-67.

- Orpinas, P. & Kelder, S. (1995). Students for Peace Project: Second student evaluation. Unpublished manuscript. Houston, TX: University of Texas Health Science Center at Houston, School of Public Health.
- Osmondson, E.L. (2000). *An evaluation of Second Step: A social skills training curriculum for violence prevention*. Unpublished doctoral dissertation, The University of West Florida.
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- 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.
- Riese, S.A. (2004). The efficacy of a violence prevention intervention with elementary school children. *Dissertation Abstracts International*, 65 (7), 2491A.
- Roberts, W., Horn, A., & Battistich, V. (1995, April). Assessing students' and teachers' sense of the school as a caring community. *Paper presented at the meeting of the American Educational Research Association*, San Francisco, CA.
- Rubin, D.B. (1991). EM and beyond. *Psychometrika*, 56, 214-254.
- Schick, A., & Cierpka, M. (2005). Faustlos: An evaluation of a curriculum to prevent violence in elementary schools. *Applied and Preventive Psychology*, 11, 157-165.

- Shure, M.B. & Spivack, G. (1982). Interpersonal problem-solving in young children: A cognitive approach to prevention. *American Journal of Community Psychology*. 10, 341-356.
- Silverman, A. (2005, May). Implementing social and character development programs in elementary schools: Challenges and baseline data from a violence prevention and social competency program in Maryland. Poster presented at the 13th Annual Meeting of the Society for Prevention Research, Washington, D.C.
- Silverman, A. & Gottfredson, G. (2005, October). A multi-method approach to evaluating the implementation of a social and character education program in a sample of Maryland elementary schools. Presentation at the Joint Canadian Evaluation Society/American Evaluation Conference, Toronto.
- Solomon, D., Battistich, V., Watson, M., Schaps, E., & Lewis, C. (2000). A six-district study of educational change: Direct and mediating effects of the Child Development Project. *Social Psychology of Education*, 4, 3-51.
- Stern, M.B. (1999). Unique Minds School Program. Unpublished manual, Ackerman Institute for the Family, 149 East 78th Street, New York, NY 10021.
- Valente, E., & Dodge, K.A. (1997). Evaluation of prevention programs for children. In R.P Weissberg, T.P. Gullotta, R.L Hampton, B.A. Ryan & G.R. Adams (Eds). *Establishing Preventive Services* (pp.183-218). Thousand Oaks, CA: Sage Publications.

- Weissberg, R. P., & Elias, M. J. (1993). Enhancing young people's social competence and health behavior: An important challenge for educators, scientists, policymakers, and funders. *Applied and Preventive Psychology, 2*, 179-190.
- Weissberg, R.P., Gesten, E.L., Carnrike, C.L., Toro, P.A., Rapin, BD, Davidson, E., et al. (1981). Social Problem-Solving Skills training: A competence-building intervention with second- to fourth- grade children. *American Journal of Community Psychology, 9*, 411-423.
- Wilson, S.J. & Lipsey, M.W. (2005). Revised report for the National Institute of Justice school violence prevention research planning meeting. Washington, DC.
- Wilson, S. J., & Lipsey, M. W. (2006). The effectiveness of school-based violence prevention programs for reducing disruptive and aggressive behavior: A meta-analysis. *International Journal on Violence and Schools, 1*, 38-50.
- Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33(2S)*, 130-143.
- Wilson, S.J., Lipsey, M.W., & Derzon, J.H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136-149.
- Wilson, D.B., Gottfredson, D.C., & Najaka, S.S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology, 17*, 247- 272.



