ABSTRACT

Title: FIRST-YEAR SELF-REPORT OUTCOMES
                                            OF A CHARACTER EDUCATION
                                            EXPERIMENT WITH ELEMENTARY
                                            STUDENTS

                                            Elise T. Harak, Master of Arts, 2006

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

Instruction in social competence in elementary grades may provide a means of preventing later problem behaviors. Previous studies indicate that school-based social competency curricula sometimes lead to decreases in problem behaviors. This randomized, controlled trial measures the efficacy of the *Second Step* program in twelve schools. The assessment of efficacy is based on 11 scales in a pre- and post-test, student self-report survey. These scales measure: (a) outcomes directly targeted by the curriculum, (b) school climate, and (c) other related outcomes. Results for students in six intervention schools are compared to students in six randomly equivalent control schools. After the first of three years of intervention, there is a statistically significant main effect for treatment on Engagement in Learning, an interaction of treatment with individual characteristics on Sense of School as a Community and Self-Restraint, and positive but not significant effect sizes on Self-Restraint and Hostile Attribution Bias.

First-year Self-Report Outcomes of a Character Education Experiment with

Elementary Students

By

Elise T. Harak.

Thesis 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
Master of Arts
2006

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

Acknowledgements

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.

Table of Contents

List of Tables

Chapter 1: Introduction

The idea that character traits are related to habits that can be taught originated decades before social scientists began investigating the effectiveness of character education in schools (Hartshorne, May & Shuttleworth, 1930). The notion that moral knowledge is linked to individual differences such as age and gender as well as environmental influences in the school and home laid the groundwork for interventions to enhance students' social knowledge within the schools. The type of social knowledge and character of concern to Hartshorne, May, and Shuttleworth was related to personality traits (e.g., honesty), not concrete skills.

Spivack, Platt, and Shure (1976) extended these ideas beyond traits to specific skills by addressing the "interpersonal cognitive problem-solving skills" (p.5) needed by individuals to get along with others effectively. Among these skills are: (a) awareness of problems one may face, (b) the ability to generate a variety of solutions and communicate the steps involved in the problem-solving process, (c) thinking about consequences, and (d) an understanding of how events relate to actions--all of which are addressed by modern day social competence programs.

For many years, researchers treated different problem behaviors (aggression, stealing, substance use, school dropout, etc.) as relatively distinct areas for inquiry. In more recent years, researchers have recognized the interrelations among problem behaviors. Contemporary research focuses on social competency development as a means of preventing many problem behaviors that are viewed as having a common set of causes (Newcomb, Abbott, Catalano, Hawkins, Battin-Pearson, & Hill, 2002).

Weissberg, Kuster, and Gullotta (1997) highlighted the need for our nation to develop effective prevention services to improve children's lives by decreasing drug use and abuse, sexual behavior that results in STDs or pregnancy, violence, and mental health problems such as depression and suicide (and improving nutrition and academic performance). They also recommended that preventive services should be applied in elementary school. Elementary grades prevention programs of known effectiveness that are well implemented are required to pursue this recommendation.

Historically the term "character" was used in relation to social knowledge and morality, but more recent social science research has generally shifted to social competence or prevention. Character education, social competence, and prevention programs all aim to increase positive traits such as helpfulness, caring, trustworthiness, and responsibility in order to decrease negative social behaviors. The following sections review literature about school-based prevention programs in general and social competence programs in elementary schools specifically. While "character education" is not often mentioned, these programs are relevant because they address skills required for the display of character. Both prevention and character programs target increasing prosocial skills and decreasing negative, antisocial behaviors. The linkage between the ideas of "character education" and the development of social and emotional competencies underlies the Social and Character Development Research Program initiated by the Institute of Education Sciences (2003) in its response to a Congressional mandate to conduct research on character education.

Meta- Analyses of School-Based Prevention Programs

Meta-analysts have examined the effectiveness of school-based prevention programs with the general population in samples ranging in age from pre-school through high school. A summary of findings from meta-analyses are discussed, starting with the most relevant and thorough studies. S.J. Wilson and Lipsey (2005) did a meta-analysis of 219 studies about prevention programs targeting aggression. The majority of these studies focused on elementary and middle schools. The meta-analysts referred to interventions as universal; selected, or those targeting at-risk individuals; and indicated, those targeting an even smaller population (Mrazek & Haggerty, 1994). Only about one-quarter of the studies involved universal prevention programs, most involved selected, and some were indicated. Wilson and Lipsey found mean effect sizes ranging from 0.20-0.30 for negative behaviors such as aggression, attention, problem behaviors, and hostility/rebellion. The interventions had mean effect sizes of about 0.20 on classroom participation and performance. The largest effects were on social skills such as communication, social problem solving, and conflict resolution (MES = 0.40). Measures in the studies included teacher reports, self-reports, and observations.

Programs that targeted selected or indicated populations were more effective than universal programs (MES= 0.30 for selected and indicated), although the 95% confidence intervals for all types overlapped. The universal programs showed an average effect size of 0.18. None of the program types had confidence intervals that overlapped with an effect size of zero.

Universal programs studied by Wilson and Lipsey (2005) involved four training approaches. In order of descending popularity, they were: (a) cognitive, (b) social skills, (c) behavioral, and (d) counseling. It should be noted that only two studies used counseling and these were reality therapy-type groups conducted in the classroom setting. Training approaches affected outcomes differently depending on the type of outcome measure used. On self- and teacher- reports (or subjective measures), effect sizes were 0.16 for behavioral programs, 0.15 for cognitive, 0.10 for social skills, and 0.43 for counseling. On archival measures (i.e., – school records) and observations (or objective measures), effect sizes were 0.33 for cognitive programs, 0.30 for social skills, and 0.16 for counseling. When entered as dummy variables into the regression models, the training approaches did not explain any unique variance when other variables (i.e., attrition, age, level of risk) were controlled.

In final regression models presented in Wilson and Lipsey (2005), research design factors for individual studies accounted for 36% of the variance in size of the effects. Objective measures and the inclusion of high risk students were associated with larger effect size estimates when other variables were controlled for. Rate of attrition, age of participants, the risk by age interaction, and the length of time implemented were negatively associated with effect size. Effects were larger across all programs when there were no implementation problems. A weakness of this finding was that Wilson and Lipsey based the classification of "having" versus "not having" implementation problems on whether the researcher reported problems. They

did not require a formal measure of implementation. Therefore, conclusions about implementation should be interpreted with caution.

Durlak and Wells (1997) examined 177 studies, 129 of which were in school settings. Overall, of these 129 studies, effect sizes for aggression were approximately normally distributed with an average ES of 0.34 and a range of -0.45 to 2.36 (95% confidence interval or CI). More specifically, programs that targeted environmental change had a moderate mean effect size of 0.35. Examples of these programs include: (a) concurrently focusing on classroom management, increasing the number of positive exchanges between teachers and students, and delivering social skills training, (b) a change in the school climate through shifting the school's organization and rules, and (c) problem-solving training. Response to social skills training differed across age groups. With students aged two to seven, affective education and interpersonal programs showed large effects of 0.70 (0.49-0.91, 95% CI) and 0.93 (0.66-1.19, 95% CI), respectively. For students aged seven to eleven, there were small to moderate effects. Affective education programs showed mean effects of 0.24 (0.18-0.31, 95% CI); interpersonal problem solving programs showed mean effects of 0.36 (0.23-0.48, 95% CI).

S.J. Wilson, Lipsey and Derzon (2003) completed a meta-analysis of 221 studies examining the effects of school-based interventions on aggression. Unlike the Wilson and Lipsey (2005) article, program format was not considered in assessing the effectiveness of the prevention programs reviewed. Within the 221 studies reviewed, there were 522 treatment and control groups. They found that randomized studies yielded an average effect size of 0.35; nonrandomized had an effect of 0.13; and

studies with only one-group and comparing two treatments (without a control group) showed effects of 0.29. It is possible that a more rigorous design did not cause larger effects to emerge; rather, those completing randomized trials may have gone to greater lengths to ensure that interventions were implemented with integrity. Prevention programs used with preschool and high school students had the most pronounced influence, with mean effect sizes of 0.33 and 0.37, respectively. Programs delivered to elementary and middle school students showed a lower mean effect size of 0.17. The most efficacious programs were behavioral/classroom management (MES= 0.43), followed by counseling (MES= 0.41), academic (MES= 0.28), and social competence programs that used a cognitive-behavioral approach (MES= 0.22). Across all of the studies and types of programs, the detection of statistically significant effects was negatively influenced by small sample sizes, high rates of attrition, and low implementation integrity. Studies that used multiple assessments and measures closely linked to the intervention showed larger effects.

D.B. Wilson, Gottfredson and Najaka (2001) did not find counseling interventions to be as efficacious as the S.J. Wilson et al. meta-analysis did. On the contrary, most of the effect sizes found were negative. However, in their classification of studies, cognitive-behavioral counseling interventions were classified separately from other counseling interventions (G.D. Gottfredson, Gottfredson, Czeh, Cantor, Crosse, & Hantman 2000). That is, cognitive-behavioral interventions were not called counseling. S.J. Wilson, Lipsey and Derzon (2003) did not have counseling broken down in such a way; therefore the positive effects they reported for counseling likely reflects the positive effects of cognitive-behavioral counseling. D.B Wilson,

Gottfredson and Najaka found environmental and cognitive-behavioral approaches to be effective.

Beelman, Pfingsten and Losel (1994) meta-analyzed 49 quasi-experimental and experimental studies that related to the effects of behavioral and cognitive-social competence training with 3- to 15-year-old children. They presented the results of treatment in terms of the particular approach (behavioral, social problem solving and self-control) and the modality (mono- or multi-modal). They found equally large effect sizes (0.77 and 0.79) for mono- and multi-modal approaches on social-cognitive skills. Effects on social interaction skills, social adjustment, and cognitions relating to the self were lower. Multimodal approaches outperformed mono-modal approaches on social interaction skills and social adjustment.

Beelman et al. also reported the influence of student characteristics on outcomes. They found that 3- to 5- year olds showed the most pronounced changes with an effect size of 1.12, followed by 9- to 11-year olds (MES= 0.91), 6- to 8-year olds (MES= 0.55), and 12- to 15- year olds (MES = 0.52). Finally, students who were characterized as socially deprived or had effected by adverse life events were benefited the most (MES= 1.06), followed by students without these risk factors and did not display externalizing, internalizing, or learning problems (MES = 0.75), and by those with externalizing or internalizing symptoms (MES = 0.67). It should be noted that all effect sizes are at least moderate and are larger than those reported by other meta-analysts. One possible reason this may be is that unpublished studies were not included in this meta-analysis. Perhaps studies that are published tend to show treatment effects, while unpublished studies may not, which would inflate these

findings as compared to other meta-analyses that did include unpublished studies (Rosenthal, 1979). (See Appendix A for a table summarizing the meta-analysis results).

Overall, the findings in all of the meta-analyses indicate that prevention programs targeting the education of social competence or character skills moderately affect negative behaviors such as aggression, social skills, and academic performance and participation. The effects vary based on the program format, with at-risk students benefiting the most, and the program approach. Cognitive and behavioral approaches tend to yield the highest effects. Finally, the research design is also important; program evaluations using a randomized, controlled trial result in larger effects than quasi-experimental designs.

Effectiveness of the Social Competence Programs in Elementary Schools

The following paragraphs review a few published evaluations of commonly used social competence programs. Although a wide variety of programs are used in elementary schools to address problem behaviors and social competence, few have been subjected to careful research. Evaluations are reviewed in order of rigor of the research, starting with the most well-researched program.

Fast Track Program. Among the well-researched programs is the Fast Track Prevention Program which utilized an adaptation of the Promoting Alternative Thinking Strategies, or PATHS, curriculum (Conduct Problem Prevention Research Group, 2002b; CPPRG, 1999b). This program was evaluated over the course of three years and resulted in a number of publications. The Fast Track program is a problem-solving skills curriculum that is used as a means to prevent problem behaviors. It

targets both the universal population of a school and a selected population that is at-risk of problem behavior, as determined through a screening process (CPPRG, 1999b). It was included in the Wilson and Lipsey (2005) meta-analysis.

At the universal level, teachers delivered an adapted version of the PATHS curriculum for 20 to 30 minutes on 2 to 3 days per week for three years. PATHS was originally designed for students with special needs, so the researchers adapted the lessons to address all students. About 80% of the lessons used for this multi-year evaluation came from the PATHS curriculum. The lessons focused on vocabulary related to affect and emotions and skills relating to social competence, prosocial behavior, self-control, and problem solving (CPPRG, 1999b).

Educational Coordinators (ECs) who were experienced teachers hired for the purpose of this project and had backgrounds in education, special education or counselor education helped teachers deliver the intervention effectively (CPPRG, 2002b). The ECs helped deliver the lessons through modeling and team teaching in the classrooms. They also observed curriculum delivery and provided feedback to the teachers. On average, this took between 1 and 1.5 hours per week. They also met regularly with the teachers for behavioral consultation (CPPRG, 1999b).

At the selected level, parents received weekly parenting classes and students took part in social skills groups and academic tutoring which were led by ECs and family coordinators. This was in addition to the universal intervention. For 30-minutes, parents and students worked together to practice their new skills. Families were given $15 for attending these sessions. Home visits and phone contacts were also conducted on a weekly basis (CPPRG, 1999a).

The researchers expected that by integrating both a universal and targeted intervention, the effects would be additive (CPPRG, 1999b). They also expected that the goal of decreasing problem behaviors in adolescence would be met since this served as an intensively delivered, early intervention (CPPRG, 2000). The researchers expected to show that the program was effective at decreasing negative behaviors and increasing positive, prosocial behaviors (CPPRG, 2000).

The CPPRG conducted a randomized evaluation of the Fast Track Program in three urban areas with low SES, minority populations and one rural area with a predominantly White and low SES population over the course of three years (CPPRG, 2000). In total, 378 classrooms across three grades participated, with 198 intervention classrooms and 180 controls. There were 7,560 student participants total (CPPRG1999b). Of those children, 891 were classified as high risk (CPPRG, 1999a).

The research group used both general linear modeling (GLM) and hierarchical linear modeling (HLM) to analyze the data. They conducted HLM analyses twice to assess the outcomes of the universal program, both with and without high-risk students included in the classroom means (because this group received the additional, targeted interventions). The researchers reported that there were no statistically significant differences in the model with or without the high-risk population, but chose to present the data without this group in order to conduct a more conservative analysis of the universal intervention (CPPRG, 1999b).

With HLM, the researchers found high intra-class correlations (ICCs) for the teacher ratings, which they hypothesized decreased the power of the design to detect the effects of the program. They contrasted this finding with the significant results

that the GLM procedure found from the teacher ratings, making the point that it is plausible that the HLM analysis did not find effects because it over-corrected for error. However, it is likely that the opposite is true; the GLM procedure likely underestimated error and overestimated the effects. Also, the researchers analyzed the outcomes in HLM as a two-level model with the classroom as the second level because they believed that having schools at the third level would not add extra variance to the model. They failed to look at school- level effects as random effects, even though schools were the unit of randomization, and thus they underestimated error components.

Based on the GLM analyses, the researchers reported that the intervention classrooms showed less aggression and improved peer relations (ES = .23 and .27, respectively) compared to the control classrooms according to student and observer ratings, but not teacher ratings. No differential effectiveness for gender and ethnic groups was found. The intervention was equally effective in the urban and rural settings. The effectiveness of the program varied with the level of implementation as well as the teachers' understanding of the curriculum. Teachers that reported a higher dosage of implementation and understanding of the curriculum rated students as less aggressive and had more positive classrooms, as determined by observations. The average teacher reported implementing 48 out of 57 lessons.

The Unique Minds School Program. A less researched program is the Unique Minds School Program or UMSP (Stern, 1999). This program "targets student proximal cognitive-social-emotional competencies . . . and distal academic learning" (Linares et al., 2005, p.406). This curriculum utilizes 30-minute, manualized lessons

and integrates conversations between students and teachers about ways to solve problems effectively as a way of learning. It is characterized as a universal prevention program.

The Linares et al. (2005) study was a quasi-experimental evaluation with a non-equivalent comparison group. There were 57 intervention and 62 control students in 13 fourth grade classrooms in New York City. About one-third of the students were White; the rest were Hispanic, Asian and Arabic. The investigators found that students who received the intervention for two years felt more socially efficacious as compared to those students in a "comparable public school" (p. 407, ES = .55); had more knowledge of prosocial, problem-solving strategies according to interviews with observers (ES = 1.01); were rated as more socially and emotionally competent by their teachers (ES = .48); and showed statistically significant improvements in math but not reading (ES = .42 and .24, respectively). The control students showed an increase in problem behaviors, while the treatment students showed a decrease. Observations of social climate, students' behavior, and cooperative communication were made by observers who were not told which schools were receiving the UMSP curriculum. According to these observations, classroom climate did not change as a result of the treatment.

Although the findings of this study seem promising, a few limitations should be considered. First, only 119 students attending two schools composed the sample, which brings into question how representative the sample is. More important, treatment assignment was not random; the school that had already implemented the curriculum was considered the treatment group and the non-equivalent comparison

group was chosen based on proximity to the treatment school, the population served (similar racial and SES composition), number of suspensions, and reading achievement scores, making the selection threat to internal validity highly plausible. The researchers did not allow for classroom variability in their models of the outcomes, making the underestimation of sampling error a threat to statistical conclusion validity.

Although the students' first experience with the curriculum occurred during this study, teachers had already taught the curriculum for 1 to 2 years prior to the beginning of the study. It is likely that the two-year exposure to the curriculum prior to the evaluation led to effects emerging earlier than they would have if the teachers did not have this experience. In other words, in other schools, the effects of this program may not emerge in the two years that emerged in this study. Also, increased student social self-efficacy does not guarantee an increase in prosocial behaviors. Furthermore, observers who were blind to treatment status did not see a change in the classroom climate, which may be a more valid source of information about increased cognitive-social-emotional competencies than student and teacher reports. On the positive side, this study involved a two-year curriculum with a diverse sample, multiple assessments across different raters and two measures of implementation fidelity.

The Peace Builders Program. A third program which has been implemented and researched with elementary school aged students is the Peace Builders program; a universal, violence prevention program that targets kindergarten through fifth graders (Embry, Flannery, Vazsonyi, Powell & Atha, 1996). This program does not involve a

fixed schedule of instruction in a curriculum; rather, it involves activities that are to be woven into the teachers' daily routine (Flannery et al., 2003). A number of tools are used to ensure this integration, such as workbooks, planning guides, binders with materials, and parent materials (Embry et al., 1996).

The goal of Peace Builders is to manipulate the antecedents that elicit negative behaviors, thereby preventing misconduct from occurring (Flannery et al., 2003; Embry et al., 1996). School staff members are expected to (a) serve as prosocial models more frequently than is typical when an intervention is not present, (b) reward students for prosocial behaviors, and (c) avoid the reinforcement of negative behaviors (Flannery et al., 2003). The intervention sets forth five rules: (a) praise others, (b) avoid insulting others, (c) seek advice from wise people, (d) right your wrongs, and (e) notice when you hurt others and correct it.

Embry et al. (1996) and Flannery et al. (2003) used a randomized trial to examine the effectiveness of Peace Builders over the course of two years in four schools. Eight schools were matched and randomly assigned to treatment level. The primary variable used for matching was the geographic proximity of the schools. The researchers also "considered . . . ethnic [ity] . . . free and reduced-price lunch . . . and ESL classrooms" (p. 295). Since the first consideration was geographic proximity, the schools looked different on the latter variables. Therefore, the matching may not have adequately eliminated between school variance. The four control schools in this study had a delayed treatment status and began to implement Peace Builders halfway through the study. These schools comprise a diverse sample, with students from a variety of ethnicities, types of households, and economic status.

14

To measure the effectiveness of the program, baseline and outcome data measuring aggression, social competence, prosocial behavior, and peace-building behaviors displayed by the students were collected in the form of teacher ratings and child self-reports. Teachers also reported on the acceptability of the program and implementation integrity. Overall, teachers agreed with the program's basic philosophy and a majority reported using it on either a daily or weekly basis (Flannery et al., 2003).

Flannery et al. (2003) reported that after the first year of the study, the treatment students showed increased social competence, decreased aggression and increased peace- building behaviors. Effect sizes and information to compute them were not included in the article. After one and a half years, the delayed intervention group had implemented the program for six months, but differences between the two groups still favored the initial implementation group. After two years, the treatment group showed more social competence, less aggression and more prosocial behaviors than the treatment-delay group (with the exception of the third graders who reported less prosocial behavior). A final finding was that students who appeared more negatively at baseline showed the greatest positive change, but the researchers did not discuss a significant treatment by initial status interaction (Flannery et al., 2003).

The ISA-SPS Curriculum. The Improving Social Awareness- Social Problem Solving (ISA-SPS) curriculum extended Spivack, Platt, and Shure's (1976) interpersonal cognitive problem-solving approach. ISA-SPS teaches students skills such as impulse control and awareness; steps involved in effective problem solving;

and generalization of what the students have learned. This curriculum is delivered over two years and was designed for elementary school-aged children.

Elias et al. (1986) examined the preventive effects of this curriculum in a post-test only, quasi-experimental study with a non-equivalent comparison group. The 158 participants from four elementary schools were in fifth grade and on average, scored one year above grade level on an achievement test. There were three groups: full implementation, a delayed intervention group and a no-treatment group. Students in the delayed group received an instructional, but not application, phase. The instructional phase was made up of 20 forty-minute lessons. During the application phase, teachers helped resolve conflicts in the classroom using the skills taught and were given activities to use on a regular basis.

The researchers found that students who received training showed better adjustment in middle school and a lower frequency and intensity of conflicts with peers than the control group. Students in the full implementation group showed the largest gains. The social problem-solving training served as a mediator for improved social adjustment. Results were reported in terms of differences in means and statistical significance of chi-squared tests. In the outcome analyses, no statistical controls were applied to account for clustering effects.

As a follow-up to an implementation of the ISA-SPS curriculum, Elias et al. (1991) attempted to measure the lasting effects of ISA-SPS by working with the original sample when they were in high school. The students had not received any intervention since elementary school. Based on student self-reports, the students who had received ISA-SPS showed scores that were statistically significantly higher on

16

language arts and math. Overall school achievement was also significantly higher. Control students reported higher levels of delinquent behavior. Treatment-comparison group differences were minimal; gender exerted more control over outcomes than treatment status. It should be noted that in the earlier article, differences in school achievement and delinquency were not discussed; the differences found in this article may have existed at baseline leading to a threat to internal validity.

Summary. Findings of statistically significant outcomes for social competence programs are associated with how the evaluation is designed, the samples used, and the characteristics of the program. Studies which had the most success in detecting effects in general were experiments with large samples, low rates of attrition and using archival (i.e., school records) or observational data, as opposed to teacher or student ratings. Most support was found for effectiveness with preschool students, although small to moderate effects were found across all ages. Studies involving higher risk students tended to show larger effects than studies with students who were not at elevated risk. Likewise, programs that targeted selected or indicated populations were the most effective; universal programs only showed small effects. There was no evidence presented that implies that social competence programs were more effective for different ethnic or gender groups. Finally, the research supports the conclusion that well-implemented behavioral, cognitive and environmental approaches can be effective.

Many observers have pointed out that the problem with programs targeting the prevention of conduct problems is that no program has proven effective other than in the short-term (Wilson et al., 2003; Durlak & Wells, 1997; Valente & Dodge, 1997;

Beelman et al., 1994). As can be seen in this review, few studies evaluate the effects of programs for more than a year and even fewer studies follow-up years later. This highlights the importance of longitudinal studies assessing long-term effects of social competence programs when delivered over many years.

Another issue that Elias (1997) pointed out is that when researchers find an intervention effective, how and why it works is often not considered. In program development and evaluation, it is important to have a theory about why the intervention should work and an idea about the essential components of the intervention that do work (Rossi, Lipsey, & Freeman, 2004).

Finally, in a lot of cases, when a school is in the midst of crisis, prevention programs are not likely to be implemented (Elias, 1997). D. C. Gottfredson, Gottfredson and Skroban (1998) found that in a district-wide school reform with external support, intervention components were not delivered to all of the high-risk students targeted and the intensity of the programs was lower than intended. Elias (1997) warned that without external supports, schools may not implement programs properly, but Gottfredson et al. showed that even with external support, it is difficult to ensure implementation integrity with prevention programs in disorganized schools. G. D. Gottfredson, Jones, and Gore (2002) provide another illustration of the difficulty of implementing otherwise well-designed interventions in chaotic school environments.

The Present Research

The goal of this project is to assess the effectiveness of the Second Step program (Committee for Children, 1991), directed at preventing problem behavior by

developing children's social skills. Effects of the program on 11 constructs are considered. Seven relate directly to the program's curriculum (Aggression, Rebellious Behavior, Victimization, Acceptability of Aggression, Altruism, Empathy, and Self-Restraint), two relate to school climate (Sense of Community and Feelings of Safety at School), and two that are less directly related to the curriculum (Engagement in Learning and Hostile Attribution Bias).

Despite the widespread international use of this program (Frey, Hirschstein, & Guzzo, 2000), little rigorous research has assessed its effectiveness. Second Step resembles a number of other social competency programs in that it is a universal, student-centered, violence prevention curriculum with manualized lessons that are to be delivered by teachers one to two times a week (Frey, Hirschstein, & Guzzo, 2000). The curriculum addresses four core competencies: empathy, anger management, problem solving, and impulse control. It is very similar to all of the other prevention programs which teach social problem-solving skills. It also qualifies as a character education program, as it focuses on teaching traits such as caring, responsibility, and fairness.

Teachers are expected to aid in generalizing the students' skills by incorporating the principles taught in the lessons into everyday instruction (Frey et al., 2000). Teachers are trained in a one-day workshop at the beginning of the school year. All other school personnel are trained during a half-day workshop to promote a school-wide approach involving a common language that is reinforced by every adult in the building (Frey et al., 2000).

Only two published studies have examined the effects of this curriculum. Grossman et al. (1997) used a randomized, controlled trial to examine the changes in baseline rates of aggression, neutral behaviors and prosocial behaviors after one year of implementation of Second Step in nineteen second and third grade classrooms, across six schools. None of the classrooms had the curriculum prior to the study. The hypothesis was that students who received Second Step would show less aggression and more prosocial behaviors than the students in the control schools. Along with random assignment, a strength of this study was that assessments were done prior to the start of the curriculum, immediately after the conclusion of the curriculum and six months later. The means of assessment were teacher and parent ratings of students and direct observations of students by observers.

Like Linares et al. (2005) and Embry et al. (1996), Grossman et al. found that the control students showed an increase in aggression, while the treatment students showed a decrease and an increase in prosocial behaviors. The difference between the two groups on both outcomes was statistically significant. For example, students in the intervention classrooms started with 2.2 episodes of physical aggression per hour during lunch/recess and decreased to 1.6 and 1.5 over two data collection periods. On the other hand, the control classrooms started at 1.8, increased to 2.6, and then went back down to 1.9 per hour. The same patterns emerged for negative verbal and overall negative behaviors and in the classrooms. No treatment effects were detected by the parent and teacher ratings (S.J Wilson and Lipsey also noted a failure to find effects using parent or teacher rating. They had not included the Grossman et al. report in their review). No major increase in prosocial behaviors was detected by either the

20

observations or ratings despite the Second Step objective of directly teaching social competencies.

Some flaws in the Grossman et al. design may have contributed to its limited findings of treatment effects. First, their N of 790 students nested within 12 schools is relatively small. While the total N of students is large, power is limited by the number of schools. Also, no objective measures of social skills (empathy, impulse control, anger management) taught by the program or of implementation integrity were used, making it impossible to determine whether the intervention (a) did not increase prosocial behaviors or (b) was not adequately implemented. Next, there was only one year of intervention. It is likely that extended exposure to the program is needed to display increased social competence and decreased aggression. One question about the validity of the one finding they did show is whether the observers were blind to the study's hypotheses. If they were not, it is possible that being attached to the research team, they were sensitized to the expected effects and were more likely to report "seeing" the changes.

The second study by McMahon, Washburn, Felix, Yakin, and Childrey (2000) examined the effectiveness of Second Step after one year of implementation in two settings with preschool and kindergarten students who ranged in age from 3 to 7 years. This was a pre-test, post-test without a control group design. The settings were chosen based on their request for violence prevention services; they differed from each other in a variety of ways; and the total number of participants was 109. Whereas most studies of elementary social skills programs have focused on suburban, predominantly White, middle SES students, this sample was almost entirely minority

students, all of whom received free-lunch. Like Grossman et al., McMahon et al. used child interviews, teacher ratings, and behavioral observations, all of which were administered prior to the beginning of the curriculum and immediately after completion. The purpose of the child interviews was to assess knowledge and skills related to empathy, impulse control, problem-solving and anger management--the core components of Second Step.

The results showed that students exhibited more knowledge and skills related to the core components of the curriculum at posttest than they had at pretest and that this was particularly evident with the older children. Based on the teachers' ratings, it appeared that the students in the preschool decreased in problem behaviors while the students in the elementary school increased. The observers detected significant decreases in verbal aggression, disruptive behavior and physical aggression. The observational data showed a larger decrease in disruptive behavior for the kindergarten students than for the preschool students -- a pattern opposite the one detected by the teacher ratings. Due to the small sample size, lack of a control group, and the contradictory findings, conclusions that Second Step decreases aggression and increases social competence cannot be drawn based on this study.

Although these studies show minimal effects of Second Step, this outcome may reflect the many threats to validity in these studies. More rigorous research that accounts for these threats, allows for a longitudinal evaluation of effects, employs adequate sample sizes, and uses the proper statistical procedures is required to determine whether Second Step is effective. This is particularly true because Second Step includes many of the components known to be needed for effective interventions

(i.e., homework, role playing, feedback, and aids for generalization) and is widely used (Martin & Pear, 1999).

There have also been 15 unpublished dissertations regarding the effectiveness of Second Step with elementary-aged students. The following six dissertations all used a non-randomized, control group trial with pre- and post-tests to examine the effectiveness of the Second Step curriculum with elementary school-aged students. None of the dissertations were true experiments. (See Appendix B for a listing of dissertations not discussed here and reasons why).

Riese (2004) examined Second Step after four months with 269 third through fifth graders. Two hundred nine students served as a comparison group. Only fourth graders reported higher empathy in the treatment than control schools; all students in the treatment schools demonstrated more understanding of impulse control (a core competency targeted by Second Step) than those in the comparison school, fourth and fifth graders showed more understanding of anger management (another core competency), and only fifth graders showed a decrease in aggression and antisocial behaviors. Teachers of third and fifth grade reported increased prosocial behaviors. Selection is a threat to the internal validity of this research.

Botzer (2002) studied the effectiveness of Second Step when delivered by a social worker over the course of a school year. There were 118 treatment participants in one school and 71 comparison students in two schools. All students were in third grade. No differences were seen in discipline referrals following implementation. On teacher ratings of aggression, one control school was significantly higher than the experimental group, the other was not. Assessments of knowledge and skills taught

23

were given to the students and showed improvements after treatment, but the comparison group never received a pretest, so selection bias is an obvious threat to the internal validity of any inferences about program effectiveness.

Lillenstein (2001) examined the effectiveness of Second Step after six months of weekly lessons with 184 treatment students (in nine classrooms) and 101 comparison students (in six classrooms). Students were in kindergarten, first and second grade. No effects of the intervention were detected by parent or teacher ratings of aggression and prosocial skills or through behavioral observations completed by the researcher. Inferences about effectiveness or ineffectiveness are not warranted because selection bias (at least) is a threat to the validity of conclusions.

McDonald (2001) studied 104 treatment and 124 comparison students in two schools. The treatment students showed a decrease in the number of office referrals. Teachers in the Second Step school may have decreased their referrals for reasons other than a drop in problem behavior (e.g., a desire to please the researcher or a change in policy about referrals). The comparison group showed higher scores on reading and math achievement. Treatment students rated themselves as having more Second Step skills than the comparison group. Selection bias (at least) is a threat to the internal validity of conclusions based on this study.

Osmondson (2000) examined the effects of Second Step after four months of implementation in three treatment schools (1,850 students) and three comparison schools (1,865 students). Comparison schools were chosen because they were the only schools in the county not using the curriculum; treatment schools were chosen based on matching demographics. Teacher ratings of students in the treatment schools

showed a statistically significant decrease in aggression and impulsivity but not an increase in prosocial behaviors. Students' ratings of each other showed significant improvements in all three areas. Despite these pre-post differences, differences between students in the treatment and comparison schools were not found for all three areas. Accordingly the pre-post differences can not be taken as treatment effects.

Finally, Harris (1998) compared the effectiveness of Second Step to a peer mediation program. Seven third and fourth grade classrooms in one school were randomly assigned to either the mediation (n = 60), Second Step (n = 59), or a control group (n = 20). Second Step lessons were delivered biweekly for four months. Results differed by grade, with one grade showing that the mediation group outperformed the Second Step group on teacher and student ratings of social skills while the other grade showed the opposite. Both groups outperformed the control group. Tests of student knowledge on problem solving, students self-report of conflict resolution and number of office referrals did not reflect treatment effects; changes from pretest to posttest did not occur in any of the groups.

All of these dissertations focused on student, not classroom or school effects. Because the interventions are delivered to groups, the failure to take account of clustering of participants within groups systematically underestimates the probability of type I errors in the statistical analyses. Results were presented only in terms of traditional analysis of variance, with results judged by statistical significance testing without taking clustering into account. In addition to using more appropriate statistical methods, it would have been more informative had the results been presented in terms of effect sizes as well as (incorrect) significance levels.

Due to design flaws, all of these dissertation studies suffer from threats to validity. None provided sound rationale to rule out alternative possibilities for the results found. Also, the researchers did not measure the integrity of implementation in a systematic way. This makes it difficult to determine whether studies that did not see changes in the treatment schools were the result of implementation problems. The findings of these dissertation studies are inconclusive about effects of Second Step when used with elementary students. The current study addresses all of these concerns because it is a randomized trial that is analyzed using hierarchical data analysis (with school- level clustering accounted for) and monitors implementation integrity for each lesson taught.

The research questions for the current study are:

1- How effective is the Second Step curriculum in increasing social competence in students after one year of implementation as reflected by student self-reports?

2- Is there differential effectiveness for different ethnic and gender groups? Based on the literature, the hypotheses are that:

1- Student surveys will reflect increased social competence as a result of the intervention.

2- There will be no treatment interactions with ethnicity and gender. While there are likely to be differences between outcome scores on ethnic and gender groups, there is nothing in the literature to suggest that Second Step would have differential effectiveness for different groups.

Chapter 2: Methods

Participants

This study is part of a larger randomized, controlled trial. All first through fifth grade students and teachers in 12 elementary schools in Anne Arundel County, Maryland were included in the study. This is one of eight sites being funded by the Institute for Education Sciences (IES) to complete an evaluation on character education programs; the third graders from year one are being followed at each site as the national study. Within the local study, first and second graders did not complete a student survey. This study examines the fourth and fifth graders because they completed student surveys and were a part of the local study. The third grade cohort will be excluded due to delays in obtaining third grade data from the contractor who collected these data.

Schools were included in the study if they never had used the Second Step curriculum and if the principal agreed to participate whether randomly assigned to treatment or control condition. Schools were matched into pairs based on their size, ethnic composition, and rates of participation in the free-and-reduced-meals (FARMS) program. Two of the schools did not have a match on all of these variables; one school has 600 students in it with no other large school to match to it and 75% of the student population in one school is ethnic minorities with the closest match having 33%. Using computer-generated pseudo random numbers, one member of each pair was randomly assigned to treatment and one to control conditions.

At the beginning of the 2004-5 school year there were a total of 4,794 students enrolled and 210 teachers across all schools. By the end of the school year, there were

4,651 students in the sample. Of this sample, 3,167 students' parents or guardians gave permission for their child to participate and all but one of the teachers consented to participate. Attrition analyses were completed to determine whether individual characteristics were associated with increased odds of being in a treatment school and of receiving consent. Findings indicate that the odds that White and Asian students would be in the treatment schools or receive consent were about double the odds for African American, Hispanic, and Native American students. Also, students receiving Special Education services or reduced or free lunch were less likely to receive consent (see Appendix C for full listing of findings). Only the students in the fourth and fifth grade who completed a student self-report survey in the fall and spring are the focus of this study. The sample size varies for each scale (n = 267-11641) as some students did not complete all questions on the questionnaire. The scales which were completed at lower rates were those which appeared at the end of the survey.

The larger project will last over the course of three years, and data on parent, teacher and student ratings of student social competence and implementation integrity are being collected. Observations were collected during year two of the study. Data, including student self-report data, for third graders will also become available in time. The present investigation is limited to the student self-report data available at present.

Intervention

Second Step is a universal, classroom curriculum that targets social competencies in an effort to increase students' prosocial skills (or promote character) and to decrease antisocial and violent behaviors. It is a grade-specific, manualized

---

[1] The first seven scales have over 1,000 responses; the last four scales have about 300 responses. The final four scales appear at the end of the survey.

curriculum broken into four core competencies: empathy, problem solving, impulse control, and anger management. The number of lessons differs in each grade with 22 in grade one, 17 in grade two, 15 in grade three, 22 in grade four, and 22 in grade five. Each lesson is delivered by the classroom teacher and takes about 35 to 45 minutes. All aspects of all lessons are expected to be delivered by the teachers. Each treatment school was provided with one Second Step kit per classroom. Control schools will receive kits for each classroom (if they wish) at the end of the three-year study.

As part of the program, trainers from the Committee for Children (the developers of Second Step) trained all teachers for a full day on the program and its implementation. Support staff were expected to attend a half day training to become familiar with the concepts targeted in Second Step so that they can aid in generalization of the skills taught. Attendance was taken to ensure that all teachers and staff attended the training or a make-up. All of the teachers in the six schools attended either the training or make-up training.

Each school has a "character development team" which plans and develops character development initiatives and events. These teams exist in both the treatment and control schools for the purpose of monitoring the character education goals and programs used in the school. The control schools do not receive any guidance or help from the research team. The treatment schools' teams are invited each summer for a workshop in using the data from this project to set goals and make plans in their schools. Both treatment and control schools are required by the school district to set goals and make plans. Another focus of the treatment schools' teams is to monitor the

delivery of Second Step. School counselors are a part of this team and also attend a monthly meeting with the research staff to discuss project issues (e.g., consent and data collection schedules) and implementation feedback. Implementation checklists are filled out by the teachers after each lesson, where they indicate whether they completed each part of the lesson. These forms are tracked and processed by the research team. Feedback specific for each teacher, in each school is provided to the counselors on a bi-monthly basis in terms of both the percentage of lessons completed and proportion of elements in the curriculum that were delivered.

Feedback about the proportion of elements delivered is captured in a composite score consisting of six components of the curriculum: the use of video clips, distribution of handouts (i.e., parent letters and homework), the creation of lists with the student-generated problem solving steps for a lesson, teacher modeling of role plays, teachers asking the students to evaluate their role play, and student role playing (see example in Appendix D). The percentage of time that a teacher delivered each of the above is tabulated. When a log is not completed, it is assumed that the lesson was not completed, resulting in a more conservative estimate of the proportion of the curriculum delivered. A weakness of this method for collecting implementation data is that it relies on the teachers' assessment of their own delivery, which may not be completely accurate. There is no inter-rater reliability since teachers were not observed during implementation.

The character development teams at treatment schools are expected to discuss the feedback at their meetings, particularly if there are obstacles to implementation. The purpose of such rigorous implementation data collection is to ensure that

implementation occurs and to allow description of the level of implementation.

Implementation level will not be used as predictors since it is not experimentally

manipulated. (See Silverman, 2005 and Silverman & Gottfredson, 2005 for further

information on implementation feedback and scoring.).

At the end of the 2004-5 school year, implementation differed across the six schools.
Table 1

Second Step Implementation at the End of the 2004-5 School Year

School Number	Percentage of Logs Completed	Percentage of Curriculum Delivered
1	100	97
2	96	86
3	73	67
4	93	85
5	99	93
6	86	80

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. In general, implementation was high.

Student Survey Measure

The student self-report survey is composed of 83, four-point Likert-type

questions and 11 scales. Since this project is a part of a larger, national study funded

by IES, the student survey was designed by IES staff in collaboration with the seven

principal investigators at all sites based on measures used in previous research. The

survey was then piloted prior to baseline data collection and revised by a data

collection contractor prior to use. This survey was created for use with the third grade

cohort at each of the national sites following consensus-seeking discussions among

the investigators about the constructs that should by measured in the research. The

present 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. Therefore, the survey used with the third

graders in year one was different than the survey used with the fourth and fifth

graders

Student surveys were administered in the fall of 2004 prior to the intervention

and in the spring of 2005, after one school year of implementation. Ninety-nine

percent of consented students returned the survey. The teacher in each classroom read

each question aloud to keep all students at the same pace. Those students who had

aides in the classroom were able to receive additional help as needed. Despite this,

many students did not complete the entire survey either as a result of the length of the

survey or non-compliance. The survey was again administered in spring 2006.

Each scale on the survey was examined by internal-consistency item analyses

after data collection in the fall of 2004; the alpha reliabilities reported below are

based on these computations. In the spring of 2005, stability coefficients were

computed by generating correlations between the fall 2004 and spring 2005 scores.

Because this survey is being used as an outcome measure, it must be sensitive to

change. Excessively high correlations between the first and second administration,

which were six months apart, would indicate that this survey is an inadequate

outcome measure. The scales are as follows:

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 to spring stability correlation is 0.47. 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.

Sense of School as a Community. This scale was adapted by IES from Roberts, Horn, and Battistich (1995). Composed of six questions such as "Students at this school are willing to go out of their way to help someone" and "People care about each other in this school," the Likert-type scale for each item ranged from "agree a lot" to "disagree a lot." The alpha reliability for this scale is 0.82; the stability coefficient is 0.54. The goal of Second Step as a universal intervention is that it will create a common language of social competence in the school; if this is successful then feelings about school as a community should be enhanced.

Feelings of Safety at School. This scale was created by IES. Composed of five questions such as "Students feel afraid that someone will hurt them at school" and "I worry that someone will pick on me at school," the Likert-type scale for each item ranged from "agree a lot" to "disagree a lot." The alpha reliability for this scale is 0.76; the stability coefficient is 0.45. If Second Step is successful at decreasing aggression and victimization, then students should feel safer at school.

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 stability coefficient is 0.51. 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 stability coefficient is 0.50. 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 stability coefficient is 0.39. 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 stability coefficient is 0.47. Second Step addresses acting in a helpful way and doing 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 stability coefficient is 0.40. 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 stability coefficient is 0.27. 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 stability coefficient is 0.52. 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 stability coefficient is 0.30. 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.

Variables

Individual students' scale scores from the spring assessment are used as dependent variables. Individual variables are used as covariates, and the treatment indicator is a school-level variable. Zero-order correlations between each predictor

were calculated to determine the benchmark model. Independent variables tested included: (a) sex, (b) ethnicity, (c) whether a child is part of the free and reduced meal (FARM) program, (d) whether a child receives special education, (e) a composite of social economic status variables (SES) based on the zip code that the student resides in, and (f) a composite of reading and math Maryland State Assessment (MSA) scores. Individual variables which correlated significantly with a majority of the outcome variables were chosen to be in the analyses as the predictors for the benchmark model; this included: (a) sex, (b) ethnicity, (c) FARM status, (d) MSA composite, and (e) the fall score on the given survey scale. The SES composite was excluded because the correlations with the outcomes were lower than the FARMS variable and the two predictors were highly correlated with one another. Treatment status was used as the sole school-level predictor (See Appendix E for a listing of the descriptive statistics for all transformed and normalized variables). These independent variables were used for all outcome analyses.

The spring and fall survey scores (used as dependent and independent variables, respectively) were skewed in most cases. Below is a listing of 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).

Table 2

Conceptual descriptions, transformations, interpretations, and ICCs for each scale

Scale Name	Conceptual Description	Transformations and Interpretations	ICC
Engagement in Learning	Assesses students' orientation to learning tasks.	The raw data was inversed (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought closer to zero but does not look normal. Lower inverse log score indicate greater engagement.	.02
Sense of School as a Community	Assesses the extent to which students feel that people care about each other and treat each other with respect in the school.	The raw data was inversed (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought close to zero. Lower inverse log score indicate greater sense of community.	.13
Feelings of Safety of School	Measures the extent to which students feel safe at school and do not worry about being victimized.	The raw data was inversed (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought closer to zero but does not look normal. Lower inverse log score indicate greater feelings of safety.	.08
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 inversed (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought close to zero. Lower inverse log score indicate greater empathy.	.03
Self-Restraint	Measures the extent to which students report that they have the social and emotional skills to control themselves in challenging social situations.	The raw data was inversed (4-the score), transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought closer to zero but does not look normal. Lower inverse log score indicate greater restraint.	.03
Hostile Attribution	Assesses the tendency of the average student	Only spring data needed to be normalized and was transformed using the natural log, and	.06

Bias	to perceive hostile or unfriendly intentions of others in ambiguous social situations.	standardized using a z-score transformation. Fall data were just standardized with a z-score transformation. Skewness was decreased and brought close to zero. In both cases, increased scores indicate higher levels of attribution bias.	
Altruism	Assesses the extent to which the average student helps others who are hurt or being picked on, cheering others up, or intervening to halt aggression among others.	Fall and spring was only standardized with z-score (no transformation needed). Increased scores indicate higher levels of altruism.	.01
Aggression	Assesses the extent to which the average student engages in 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 and brought closer to zero but does not look normal. Increased scores indicate higher levels of aggression.	.03
Rebellious Behavior	Measures student reports of the frequency with which they engage in rebellious or 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 decreased and brought close to zero. Increased scores indicate higher levels of rebellious behavior.	.03
Victimization	Indicates the extent to which students report being the victims of teasing, name calling, threats, or other forms of physical or social aggression.	The raw data was transformed using the natural log, and standardized using a z-score transformation. Skewness was decreased and brought close to zero. Increased scores indicate higher levels of victimization.	.03
Acceptability of	Measures normative beliefs about	The raw data was transformed using the natural log, and standardized using a z-score	.02

| Aggression | aggression. This scale is a gauge of norms allowing interpersonal aggression such as yelling at or making fun of others and saying bad things. | transformation. Skewness was decreased and brought close to zero. Increased scores indicate higher levels of acceptability of aggression. |

Note: ICC = intraclass correlation.

Analyses

The number of schools assigned to conditions, the degree of clustering of students within school, and the extent to which it is possible to model between and within-school variability all influence power. Following procedures described by Schochet (2005), minimum detectable effect sizes were computed for two-tailed tests with $\alpha = 0.05$ for six matched pairs of schools (12 schools) for a variety of intraclass correlation (ICC) values. This was necessary since there are 11 different scales with ICCs ranging from 0.01 to 0.13. See Table 3 for the minimally detectable effect (MDE) sizes with a power of .8.

Table 3.

Minimum Detectable Effect Sizes (MDE)

Intraclass Correlation	Rb2	Rw2	MDE
$\rho = 0.01$	0.6	0.5	0.05
$\rho = 0.02$	0.6	0.5	0.06
$\rho = 0.03$	0.6	0.5	0.07
$\rho = 0.04$	0.6	0.5	0.08
$\rho = 0.05$	0.6	0.5	0.09
$\rho = 0.06$	0.6	0.5	0.10
$\rho = 0.07$	0.6	0.5	0.10
$\rho = 0.08$	0.6	0.5	0.11
$\rho = 0.09$	0.6	0.5	0.11
$\rho = 0.10$	0.6	0.5	0.12
$\rho = 0.11$	0.6	0.5	0.13
$\rho = 0.12$	0.6	0.5	0.13
$\rho = 0.13$	0.6	0.5	0.14

Note: Rb2 = between school variance accounted for by the statistical model. Rw2= within school variance accounted for by the statistical model.

Because schools, and not classrooms or individuals, were assigned to treatment, students within school are not independent. Instead, students are nested within schools. Therefore, data were analyzed using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) with the individual at level one and the school at level two. This ensures that the statistical significance of effects are not overestimated due to an underestimation of sampling error, and will take school into account in predicting the outcomes. Classrooms were not treated as a level because schools (not classrooms) were randomized and the treatment indicator that could be used for classrooms within schools would be the same for all classrooms in a school.

Chapter 3: Results

There are 11 HLM equations in total; one for each student survey scale with the spring score as a dependent variable. Following the advice of Raudenbush and Bryk (2002), all individual predictors were first analyzed with group-mean centering and treated as having random slopes across schools. Significance tests were conducted for random variance components. In subsequent models, all variables were grand-mean centered, and the slopes of variables with significant variance components were freed and all others were fixed. All analyses had the same five individual-level predictors (sex, ethnicity, FARM, MSA, and fall baseline score on the dependent variable) and the same school-level predictor (treatment). Non-significant predictors were retained in all models. The level-2 coefficient for treatment in the final estimation of fixed effects on intercept in the level-1 equation indicates the size of the treatment effect.

The equation for the Engagement in Learning model is as follows:

Level One (Individual)

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

Level Two (School)

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

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

β5j= γ50

In the above equations, Y_{ij} represents the spring student survey engagement scale, $\beta_{1j}(X_{ij} - X_{ij}..)$ represents the grand mean centered fall baseline engagement score and its regression weight, $\beta_{2j}(X_{ij} - X_{ij}..)$ represents grand mean centered gender and its regression weight, $\beta_{3j}(X_{ij} - X_{ij}..)$ represents grand mean centered ethnicity and its regression weight, $\beta_{4j}(X_{ij} - X_{ij}..)$ represents grand mean centered FARMS status and its regression weight, and $\beta_{5j}(X_{ij} - X_{ij}..)$ represents the grand mean centered MSA composite 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, and γ_{01} represents the treatment effect (W_{ij} being the treatment indicator). Because there were no significantly randomly varying slopes in the model student engagement, there are no error terms for the individual predictors.

The test for main effects of treatment was significant for Engagement in Learning with an effect size of -0.15 (SE = 0.03, p = 0.05). Table 4 shows the findings for the Engagement model.

Table 4

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

Fixed effects				
Variable	Coefficient [a]	SE	df	p
Intercept	-.02	.03	10	.58
Treatment status (0 = control, 1 = treatment)	-.15	.06	10	.05*
Individual sex (0 = female, 1 = male)	.07	.07	1078	.19
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.10	.10	1078	.34
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.01	.06	1078	.89
Individual fall Engagement score	.46	.03	1078	.00 *
MSA composite	-.06	.03	1078	.08
Random effects				
Variance component	Variance	df	$\chi 2$	p
Level-1 error	.88	--	--	--
Level-2 intercept	.06	10	14.30	.16
Proportion of variance explained				
Proportion of tau explained	.86			
Proportion of sigma-squared explained	.20			
Net ICC	.00			
Proportion of between school variance explained	1.00			

Note. SE = standard error.
[a]Scores are the inverse log of raw data and therefore a negative coefficient indicates an increase in engagement; a positive coefficient indicates a decrease in engagement.

The effect sizes for Feelings of Safety at School, Empathy, Aggression, Rebellious Behavior, Victimization, and Acceptability of Aggression are close to zero, with a range of ES = -.05 to .05. It should be noted that the n decreased substantially on the final four sections of the survey: (a) Aggression, (b) Rebellious Behavior, (c) Victimization, and (d) Acceptability of Aggression. It is suspected that length of the survey and non-compliance to respond played a role in these decreases. Data were explored to determine whether missing data were confined to some schools or classrooms, but the n dropped substantially for these scales in all classrooms.

The effect sizes for treatment on Self-Restraint and Hostile Attribution are small and not statistically significant but in the desired direction; they are -0.08 (SE =

0.08, p = 0.37) and -0.09 (SE = 0.11, p = 0.44) respectively. Students in the treatment schools reported being about one-tenth of a standard deviation more restrained and interpret ambiguous situations as one-tenth of a standard deviation less hostile than students in the control schools. (See Appendix F for tables of the outcomes on these variables) The effect sizes for sense of School as a Community and Altruism also approach one-tenth of a standard deviation, but not in the desired direction; they are 0.07 (SE = 0.15, p = 0.65) and -0.08 (SE = 0.06, p = 0.28) respectively.

A second purpose of this study is to determine whether treatment is differentially effective for different ethnic or gender groups. This requires that these variables have randomly varying slopes to model an interaction. When a slope is randomly varying (unlike in the case of the Engagement in Learning model), the level-2 coefficient for treatment in the final estimation of fixed effects on the slope for groups in the level-1 equation indicates a treatment-by-group interaction. None of the scales showed a treatment interaction with gender or ethnicity, but two interactions of treatment with other individual predictors in the benchmark models were found.

On Sense of School as a Community, there is a significant interaction between treatment and FARM status. In treatment schools, students receiving reduced lunch (controlling for the other individual predictors) increased about one-third of a standard deviation on Sense of Community in the spring as compared to their peers with the same FARM status in the control schools. This effect is doubled for students receiving free lunch. Students receiving subsidized meals in the treatment schools are more likely to feel that people are more respectful and caring in their school than

FARM students in the control schools, thus making the schools more equitable. FARM students in the treatment schools also find the school more respectful than students in the same school who do not receive subsidized meals. The only significant individual-level predictor of Sense of Community was fall baseline score. For every standard deviation increase in the fall, kids find the school to be one-half of a standard deviation more respectful and caring in the spring. This model accounts for one-third of the between school variance. It is possible that other school-level predictors would account for the remaining 9% of variance between schools.

On Self-Restraint, there is a treatment interaction with individual fall baseline score. Students in the treatment schools with one standard deviation below the average Self-Restraint score in the twelve schools have almost one-fifth of a standard deviation higher rating of Self-Restraint in the spring than their peers in the control schools with the same baseline score. Again, the treatment schools are more equitable because the baseline score is less predictive of the spring outcome and there is less differentiation between students based on their baseline scores. The significant individual predictors are sex, fall baseline score, and MSA score. Male students report being almost one-fifth of a standard deviation less restrained than female students. For every standard deviation increase in the fall, students report one-half of a standard deviation increase on Self-Restraint in the spring. As students increase one standard deviation on their MSA composite score, their reports of restraint also increased one-tenth of a standard deviation. This model accounts for one-third of the between school variance on the Self-Restraint scale, leaving 2% of unexplained variance between schools.

46

Because attrition and beyond chance difference in the characteristics of students in the treatment and control conditions were problems, sensitivity analyses were conducted to determine whether the outcome analyses would be the same with a complete sample or controlling for student characteristics. Two analyses were used; analyses using weights based on school, grade, ethnicity, and FARM status (Graham, Cumsille, & Elek-Fisk 2003) and imputation using an EM algorithm (Rubin, 1991). The effect sizes for all three analyses shown in Table 5 were in the same direction, but the sensitivity analyses were more conservative than the benchmark models. The main treatment effect on Engagement was found in both sensitivity analyses. In both cases, there was still interaction of treatment with FARMS status for Sense of School as a Community. Only the weighted analysis found the interaction of treatment with baseline score on Self-Restraint. In the imputation analysis, the test for a randomly varying slope was not significant in any model, and no slope could be modeled.

Table 5
Effect Size on benchmark and sensitivity analyses

Outcome	Benchmark Effect Size	Weighted Effect Size[a]	Imputed Effect Size
Treatment effect on Engagement in Learning	-.15*	-.12*	-.16*
Treatment effect on Sense of School as a Community	.07	.09	.06
Treatment interaction with FARM status on Sense of School as a Community	-.33*	-.30*	-.24*
Treatment effect on Feelings of Safety at School	.01	.03	.01
Treatment effect on Empathy	.05	.07	.05
Treatment interaction with Ethnicity on Empathy	.06	.06	.04
Treatment interaction with FARM status on Empathy	-.05	-.04	-.04
Treatment interaction with fall baseline score on Empathy	N/A[b]	N/A[b]	-.04
Treatment effect on Self-Restraint	-.08	-.06	-.08
Treatment interaction with fall baseline score on Self-Restraint	-.14*	-.13*	-.15*
Treatment interaction with FARM status on Self-Restraint	N/A[b]	N/A[b]	-.21
Treatment effect on Hostile Attribution Bias	-.09	-.09	-.10
Treatment interaction with Ethnicity on Hostile Attribution Bias	N/A[b]	.38	N/A[b]
Treatment effect on Altruism	-.08	-.07	-.05
Treatment effect on Aggression	-.05	-.06	-.01
Treatment interaction with Ethnicity on Aggression	-.31	-.24	N/A[b]
Treatment effect on Rebellious Behavior	.05	.07	.01
Treatment effect on Victimization	.03	.06	-.07
Treatment effect on Acceptability of Aggression	-.04	.00	-.02

[a]Effect sizes for the weighted analyses are based on robust standard errors. The degrees of freedom needed for this analysis is 20. These are used as an estimate.
[b]No randomly varying slope to model.

48

Chapter 4: Discussion

Second Step is an internationally used program despite the inconclusive findings supporting its effectiveness. It is likely that schools continue to use it because it includes the essential components for a successful intervention, the lessons are easy to deliver, teachers and students enjoy the content, and no similar program shows stronger support. To date, the studies about Second Step suffer from threats to validity, and therefore should be interpreted with caution. As such, a more rigorous study was needed to determine whether it is effective.

The student self-reports indicate that after one year of implementation, there is a main effect of treatment on Engagement in Learning. The interaction findings on Sense of School as a Community and Self-Restraint provide promising evidence that Second Step helps to close the gap between advantaged students and their less advantaged peers. Besides the statistically significant findings, there is also evidence of positive effects on three other scales. Although not statistically significant, it is possible that effects may increase with additional years of exposure to the treatment.

Randomization of schools to treatment and control conditions, a year-long implementation of the Second Step curriculum, and a large sample contribute to the internal validity of this study. One weakness of the present report is the use of mono-method data collection (i.e., reliance on student self-report measures). Future research will examine additional dependent measures based on teacher and parent reports and on school archival measures of achievement.

Major threats to validity are attrition and pre-existing differences between the students in the treatment and control groups. There are three problems. First, despite

random assignment there are relatively more African American, Hispanic, and Native American students in the control schools, making the treatment schools less diverse. Second, there is selective participation based on consent rates. Students in ethnic minority groups and those who come from lower income families are less likely to receive consent. Despite these problems, sensitivity analyses revealed that the conclusions about treatment effectiveness are the same for the benchmark, weighted and imputed analyses.

Third, the incomplete data on the Aggression, Rebellious Behavior, Victimization, and the Acceptability of Aggression scales imply additional selective participation. The questions on these scales were at the end of the survey; it is possible that the schools did not allow for enough time to complete the surveys (no time restraints were imposed by the researcher). Missing data might also be a function of the sensitivity of the scales; students may not have wanted to reveal their negative behaviors. It is also possible that teachers encouraged students not to respond. The data were explored to see if there was a pattern in the missing data but the problem is pervasive in all schools and classrooms.

These results imply modest positive effects of the Second Step curriculum on student self-reports. These effects were small in size and generally not statistically significant. In other words, despite evidence that the intervention was well implemented by well trained teachers who had the benefit of feedback on the level of implementation the evidence summarized here shows relatively small effects. In this respect the results resemble those of Grossman et al. (1997) who also found few positive effects of the program. On the other hand, these are in contrast to the meta-

analysis findings that suggest that universal prevention programs should have a moderate effect on a variety of outcomes.

The present results are limited only to part of the outcome data now available (i.e., to student self-report measures), to outcomes after only one year of implementation, and the number of schools is relatively small, limiting the statistical power of the analyses. Non-significant effects, modest but generally positive in direction, were found for a number of outcomes. In addition, evidence that the treatment interacted with student characteristics in a way that generally benefited students who were low in socioeconomic status, as indicated by their participation in the free and reduced meal program, and on students with lower endorsements of positive behaviors during baseline. Future analyses will examine additional outcomes and the results after two and three years of implementation.

Appendices

Appendix A. Summary of Meta-Analysis Findings

Author	Number of Studies	Overall Findings
S.J. Wilson & Lipsey (2005)	219	-Mean effect sizes of 0.20-0.30 for negative behaviors. - M_{ES} = 0.20 on classroom participation and performance. - Largest effects on social skills (M_{ES} = 0.40). - Selected/indicated programs more effective than universal programs (M_{ES}= 0.30 versus 0.18). - School records and observations detected larger effects than rating forms. -Research design of the studies accounted for 36% of the variance of effect sizes.
Durlak & Wells (1997)	177 (129 in school settings)	- Mean effect size 0.34 on aggression. - For students ages 2-7, affective education and interpersonal programs showed effects of 0.70.
S.J. Wilson, Lipsey, & Derzon (2003)	221	- Differences between effect sizes were found according to research design with randomized trials detecting the largest effects. - Programs delivered to elementary students had an effect of 0.17. - Find differential effectiveness based on the program approach: behavioral/classroom management (M_{ES}= 0.43), counseling (M_{ES}= 0.41), academic (M_{ES}= 0.28), and cognitive-behavioral social competence (M_{ES}= 0.22).
Beelman, Pfingsten and Losel (1994)	49	- Equal effect sizes (0.77 and 0.79) for mono- and multi-modal approaches on social-cognitive skills. - 9- to 11-year olds (M_{ES}= 0.91) and 6- to 8-year olds (M_{ES}= 0.55). - At-risk students benefited the most (M_{ES}= 1.06). -One reason these outcomes are higher than the others is that that unpublished studies were not included in this meta-analysis but were in the others.

Appendix B. Table of Dissertations Excluded from the Text Literature Summary

Author	Methodology	Sample Size	Comparison Group	Other Reason(s) for Exclusion
Jakob (2005)	Quasi-experimental examination of effects	34 treatment students	18 comparison students; no documentation of comparability	N/A
Lai (2001)	Ethnographic study documenting implementation	N/A	N/A	N/A
Nicolet (2004)	Quasi-experimental	54 treatment students	55 comparisons	Only completed nine of fifteen lessons
Oppitz (2002)	Descriptive survey of the activities school counselors engage in relating to violence prevention	N/A	N/A	N/A
Reed (2004)	Qualitative review	N/A	N/A	N/A
Teagarden (2002)	Differential effectiveness when delivered by a teacher, counselor, or student.	N/A	N/A	N/A
Washburn (2001)	Non-experimental examination of effects with low-income, urban youth; no comparison group used.	N/A	No comparison group	N/A
Wojtalewicz (2004)	Non-experimental examination	N/A	None	N/A
Wilke (2004)	Retrospective correlational study	N/A	N/A	N/A

Note: N/A = non-applicable; used for studies where the information was not provided or was not relevant to exclusionary criteria.

Appendix C. Attrition Information

Table C1. *Relative Odds of Being Assigned to Treatment*

Individual Characteristics	Grade 1 n = 942	Grade 2 n = 947	Grade 3 n = 980	Grade 4 n =994	Grade 5 n = 968
Gender (0 = female, 1 = male)		0.73**			
Ethnicity (0 = disadvantaged, 1 = advantaged)	1.69**	2.30**	2.47**		2.12**
SpecEd (0 = not in Special Ed, 1 = in Special Ed)				1.64**	
ESOL (0 = not in ESOL, 1 = in ESOL)					
FARM (0 = no FARM, 1 = reduced, 2 = free)					
Referrals (0 = no referrals, 1 = 1 or more)					
Suspensions (0 = no suspensions, 1 = 1 or more)				0.24*	0.08**
MSA reading (z-scored)	NA	NA	0.87**	0.67**	0.77**
MSA math (z-scored)	NA	NA	0.80*	0.74**	0.69**

Note: NA = Not available; the MSA tests are not administered at these grade levels. ESOL = English as a Second Language. FARM = free and reduced meal. Tabled values are the change in odds associated with a unit increase in the characteristic listed.
*p< 0.10, **p < 0.05

Table C2. *Relative Odds of Receiving Consent*

Individual Characteristics	Grade 1 n = 942	Grade 2 n = 947	Grade 3 n = 980	Grade 4 n =994	Grade 5 n = 968
Gender (0 = female, 1 = male)					
Ethnicity (0 = disadvantaged, 1 = advantaged)	2.52**	1.61**	1.53*		1.87**
SpecEd (0 = not in Special Ed, 1 = in Special Ed)	0.59*				0.55**
ESOL (0 = not in ESOL, 1 = in ESOL)					
FARMs (0 = no FARMS, 1 = reduced, 2 = free)			0.73**		0.76**
Referrals (0 = no referrals, 1 = 1 or more)					0.51*
Suspensions (0 = no suspensions, 1 = 1 or more)					
MSA reading (z-scored)	NA	NA	1.42**	1.31**	1.47**
MSA math (z-scored)	NA	NA	1.53**	1.28**	1.47**

Note: NA = Not available; the MSA tests are not administered at these grade levels. ESOL = English as a Second Language. FARM = free and reduced meal. Tabled values are the change in odds associated with a unit increase in the characteristic listed.
*p< 0.10, **p < 0.05

Appendix D. Example of an Implementation Log

Grade 1 Second Step Implementation Record
Unit I: Empathy Training

Date Lesson Delivered			**Lesson 1: Empathy Training-Skill Overview**		
Month	**Day**		**In conducting this lesson, did you . . .**		
⓪ Sep	⓪	⓪	Ⓨ	Ⓝ	1. Distribute take-home letter 1: Introduction to *Second Step?*
⓪ Oct	①	①	Ⓨ	Ⓝ	2. Use the photo to introduce what will be learned?
⓪ Nov	②	②	Ⓨ	Ⓝ	3. Have students generate rules for during the lessons?
⓪ Dec	③	③	Ⓨ	Ⓝ	4. Have students recite the rules?
⓪ Jan	④	④	Ⓨ	Ⓝ	5. Have a discussion about the weekend to practice using rules?
⓪ Feb	⑤	⑤	Ⓨ	Ⓝ	6. Praise students who followed the rules?
⓪ Mar	⑥	⑥	Ⓨ	Ⓝ	7. Discuss times when students can use the rules?
⓪ Apr	⑦	⑦	Ⓨ	Ⓝ	8. Do the listening game?
⓪ May	⑧	⑧	Ⓨ	Ⓝ	9. List feeling names and a face to match on poster paper?
⓪ Jun	⑨	⑨	Ⓨ	Ⓝ	10. Hang the poster with the rules on the wall?

Comments about lesson 1:

Appendix E. Descriptive Statistics for Fall Individual and School-Level Predictors and Spring 2005 Student Self Report Scales.

Variable	*N*	*M*	*SD*	Min	Max
Dependent variables (Spring *Student Survey*)					
Engagement in Learning	1283	.00	1.00	-1.62	3.48
Sense of School as a Community	1280	.00	1.00	-1.79	2.30
Feelings of Safety at School	1279	.00	1.00	-1.30	2.26
Empathy	1277	.00	1.00	-2.27	2.91
Self-Restraint	1268	.00	1.00	-1.59	2.54
Hostile Attribution Bias	1228	.00	1.00	-2.68	1.82
Altruism	1277	.00	1.00	-1.93	2.38
Aggression	690	.00	1.00	-1.31	2.09
Rebellious Behavior	560	.00	1.00	-0.91	2.76
Victimization	944	.00	1.00	-1.57	1.84
Acceptability of Aggression	533	.00	1.00	-1.37	1.94
Individual-level predictor variables (Fall *Student Survey*)					
Engagement in Learning	1235	.00	1.00	-1.63	2.81
Sense of School as a Community	1222	.00	1.00	-1.63	2.59
Feelings of Safety at School	1227	.00	1.00	-1.30	2.31
Empathy	1222	.00	1.00	-2.52	3.24
Self-Restraint	1213	.00	1.00	-1.71	2.74
Hostile Attribution Bias	1226	.00	1.00	-1.88	2.68
Altruism	1209	.00	1.00	-2.00	2.31
Aggression	590	.00	1.00	-1.18	2.53
Rebellious Behavior	570	.00	1.00	-0.82	3.44
Victimization	890	.00	1.00	-1.57	1.88
Acceptability of Aggression	542	.00	1.00	-1.40	2.13

Individual-level predictors from school
archives

Student sex (0 = female, 1 = male)	1342	0.51	0.50	.00	1.00
Student ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	1229	0.90	0.29	.00	1.00
Free and reduced meals (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	1311	0.17	0.55	.00	2.00
Maryland State Assessment (MSA) Composite (average of reading and math)	1210	0.07	0.91	-3.16	3.64
School-level variable					
Treatment Status (0 = control, 1 = treatment)	12	.50	0.52	.00	1.00

Note: Descriptive statistics are on the transformed and normalized variables.

Appendix F. HLM Outcome Tables

Table F1

Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Sense of School as a Community Made by Students, Fourth and Fifth Grade Sample

Fixed effects

Variable	Coefficient[a]	SE	df	p
Intercept	-.02	.08	10	.84
Treatment status (0 = control, 1 = treatment)	.07	.15	10	.65
Individual sex (0 = female, 1 = male)	.04	.05	1065	.46
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.04	.10	1065	.67
FARMS status slope (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.04	.06	10	.47
FARMS x Treatment Interaction	-.33	.11	10	.02*
Individual fall community score	.48	.03	1065	.00*
MSA composite slope	-.01	.03	1065	.79

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.82	--	--	--
Level-2 intercept	.25	9	66.86	.00
FARMS slope	.02	9	7.50	>.50

Proportion of variance explained

Proportion of tau explained	.53
Proportion of sigma-squared explained	.26
Net ICC	.09
Proportion of between school variance explained	.31

Note. *SE* = standard error.

[a]Scores are the inverse log of raw data and therefore a negative coefficient indicates an increase in sense of community; a positive coefficient indicates a decrease in sense of community.

Interpretation: There is no treatment effect for sense of school as a community, but there is a significant interaction between treatment and FARMS students. In treatment schools, students receiving reduced lunch increased their ratings on community about one-third of a standard deviation as compared to their peers with the same FARMS status in the control schools. This effect is doubled for students receiving free lunch; they increase two-thirds of a standard deviation on sense of school as a community as compared to their peers in the control schools. In other words, treatment schools show less differentiation between their FARMS students than the control schools showing that the treatment schools are more equitable. The only significant individual predictor is fall baseline score. For every standard deviation increase in the fall, kids find the school to be one-half of a standard deviation more respectful and caring in the spring. This model accounts for one-third of the between school variance. It is possible that other school-level predictors would account for the remaining 9% of variance between schools.

Table F2

Estimates of Fixed Effects in a Two-Level Model of Spring Ratings of Feelings of Safety at School Made by Students, Fourth and Fifth Grade Sample

Fixed effects

Variable	Coefficient [a]	SE	df	p
Intercept	-.03	.06	10	.67
Treatment status (0 = control, 1 = treatment)	.01	.12	10	.91
Individual sex (0 = female, 1 = male)	-.01	.05	1069	.84
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.21	.11	1069	.05[*]
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.03	.06	1069	.60
Individual fall feelings of safety score	.38	.03	1069	.00[*]
MSA composite	-.09	.03	1069	.01[*]

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.88	--	--	--
Level-2 intercept	.18	10	43.49	.00

Proportion of variance explained

Proportion of tau explained	.57
Proportion of sigma-squared explained	.17
Net ICC	.04
Proportion of between school variance explained	.50

Note. *SE* = standard error.

[a]Scores are the inverse log of raw data and therefore a negative coefficient indicates an increase in feelings of safety; a positive coefficient indicates a decrease in feelings of safety.

Interpretation: There is no treatment effect for feelings of safety at school. The significant individual predictors are ethnicity, fall baseline score, and MSA score. As compared to African American, Hispanic, and Native American students, White and Asian students rate the school as one-fifth of a standard deviation safer. For every standard deviation increase in the fall, kids find the school to be almost one-half of a standard deviation safer. As students increase one standard deviation on their MSA composite score, their feelings of safety is also increased almost one-tenth of a standard deviation. This model accounts for half of the between-school variance on feelings of safety at school. It is possible that school-level predictors other than treatment status would account for the remaining 4% of variance between schools.

Table F3

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

Fixed effects				
Variable	Coefficient[a]	*SE*	*df*	*p*
Intercept	-.01	.04	10	.73
Treatment status (0 = control, 1 = treatment)	.05	.08	10	.55
Individual sex (0 = female, 1 = male)	.25	.05	1063	.00[*]
Ethnicity slope (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.09	.16	10	.57
Ethnicity x Treatment interaction	.06	.31	10	.85
FARMS status slope (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.08	.10	10	.44
FARMS x Treatment interaction	-.05	.21	10	.83
Individual fall empathy score	.53	.03	1063	.00[*]
MSA composite slope	-.15	.03	1063	.11[*]
Random effects				
Variance component	Variance	*df*	χ^2	*p*
Level-1 error	.83	--	--	--
Level-2 intercept	.09	9	18.81	.03
Ethnicity slope	.36	9	14.91	.09
FARMS slope	.26	9	25.74	.00
Proportion of variance explained				
Proportion of tau explained	.66			
Proportion of sigma-squared explained	.30			
Net ICC	.01			
Proportion of between school variance explained	.67			

Note. *SE* = standard error.
[a]Scores are the inverse log of raw data and therefore a negative coefficient indicates an increase in empathy; a positive coefficient indicates a decrease in empathy.

Interpretation: There is no treatment effect for empathy. The significant individual predictors are sex, fall baseline score, and MSA score. Male students report one-fourth of a standard deviation less empathy than female students. For every standard deviation increase in the fall, students report one-half of a standard deviation increase on empathy in the spring. As students increase one standard deviation on their MSA composite score, their reports of empathy also increased one-tenth of a standard deviation. This model accounts for 67% of the between school variance on empathy. It is possible that other predictors would account for the remaining 1% of variance between schools.

Table F4

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

Fixed effects

Variable	Coefficient[a]	SE	df	P
Intercept	-.03	.04	10	.51
Treatment status (0 = control, 1 = treatment)	-.08	.08	10	.37
Individual sex (0 = female, 1 = male)	.18	.05	1047	.00*
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.20	.10	1047	.05*
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.03	.06	1047	.66
Individual fall restraint score slope	.49	.03	10	.00*
Fall score x Treatment interaction	-.14	.06	10	.03*
MSA composite	-.04	.03	1047	.23

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.84	--	--	--
Level-2 intercept	.12	10	26.34	.00
Individual fall restraint score slope	.03	10	12.65	.24

Proportion of variance explained

Proportion of tau explained	.62
Proportion of sigma-squared explained	.27
Net ICC	.02
Proportion of between school variance explained	.33

Note. *SE* = standard error.
[a]Scores are the inverse log of raw data and therefore a negative coefficient indicates an increase in restraint; a positive coefficient indicates a decrease in restraint.

Interpretation: There is no statistically significant treatment effect for self restraint; however, the effect size is approaching one-tenth of a standard deviation in the desired direction. There is also a treatment interaction with individual fall score on self restraint. Students with one standard deviation below the average self restraint score in treatment schools have almost one-fifth of a standard deviation higher rating of self restraint in the spring than their peers with the same baseline score in the control schools. In other words, the treatment schools are more equitable; the baseline score is less predictive of the spring outcome and there is less differentiation between students based on their baseline scores. The significant individual predictors are sex, ethnicity, and fall baseline score. Male students report being almost one-fifth of a standard deviation less restrained than female students. White and Asian students report being one-fifth of a standard deviation more restrained than African American, Hispanic, and Native American students. For every standard deviation increase in the fall, students report one-half of a standard deviation increase on self restraint in the spring. This model accounts for one-third of the between school variance in self-restraint leaving 2% of unexplained variance between schools.

Table F5

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

Fixed effects				
Variable	Coefficient [a]	*SE*	*df*	*P*
Intercept	-.03	.06	10	.62
Treatment status (0 = control, 1 = treatment)	-.09	.11	10	.44
Individual sex (0 = female, 1 = male)	-.04	.06	1021	.45
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	.09	.11	1021	.44
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.08	.07	1021	.24
Individual fall bias score	.32	.03	1021	.00[*]
MSA composite	-.16	.04	1021	.00[*]
Random effects				
Variance component	Variance	*df*	χ^2	*p*
Level-1 error	.93	--	--	--
Level-2 intercept	.17	10	39.17	.00
Proportion of variance explained				
Proportion of tau explained	.53			
Proportion of sigma-squared explained	.09			
Net ICC	.03			
Proportion of between school variance explained	.50			

Note. *SE* = standard error.

[a]A negative coefficient indicates a decrease in hostile attribution bias; a positive coefficient indicates an increase in hostile attribution bias.

Interpretation: There is no statistically significant treatment effect for hostile attribution bias however the effect size is approaching one-tenth of a standard deviation in the desired direction. The significant individual predictors are fall baseline score and MSA score. For every standard deviation increase in the fall students report one-third of a standard deviation increase on hostile attribution bias in the spring. As students increase one standard deviation on their MSA composite score, their reports of bias decreased almost one-fifth of a standard deviation. Half of the between school variance in hostile attribution bias was accounted for with this model, leaving 3% of the variance between schools.

Table F6

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

Fixed effects

Variable	Coefficient[a]	SE	df	P
Intercept	.01	.03	10	.87
Treatment status (0 = control, 1 = treatment)	-.08	.06	10	.28
Individual sex (0 = female, 1 = male)	-.14	.06	1051	.01[*]
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	.02	.10	1051	.82
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.06	.06	1051	.33
Individual fall altruism score	.46	.03	1051	.00[*]
MSA composite	.00	.03	1051	.94

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.89	--	--	--
Level-2 intercept	.06	10	13.93	.18

Proportion of variance explained

Proportion of tau explained	.70
Proportion of sigma-squared explained	.21
Net ICC	.00
Proportion of between school variance explained	1.00

Note. *SE* = standard error.

[a]A negative coefficient indicates a decrease in altruism; a positive coefficient indicates an increase in altruism.

Interpretation: There is no statistically significant treatment effect for altruism. The significant individual predictors are sex and fall baseline score. Males report almost one-fifth of a standard deviation less altruistic behaviors than females. For every standard deviation increase in the fall, students report almost one-half of a standard deviation increase on altruism in the spring. This model accounts for all of the between school variance on altruism.

Table F7

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

Fixed effects				
Factor	Coefficient[a]	*SE*	*df*	*p*
Intercept	.09	.06	10	.18
Treatment status (0 = control, 1 = treatment)	-.05	.12	10	.67
Individual sex (0 = female, 1 = male)	.21	.10	355	.04[*]
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.13	.20	10	.54
Ethnicity x Treatment Interaction	-.31	.39	10	.46
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.15	.10	355	.13
Individual fall aggression score	.23	.05	355	.00[*]
MSA composite	-.06	.05	355	.27
Random effects				
Variance component	Variance	*df*	χ^2	*p*
Level-1 error	.89	--	--	--
Level-2 intercept	.11	9	14.83	.10
Ethnicity slope	.42	9	18.02	.04
Proportion of variance explained				
Proportion of tau explained	.58			
Proportion of sigma-squared explained	.19			
Net ICC	.01			
Proportion of between school variance explained	.67			

Note. *SE* = standard error.
[a]A negative coefficient indicates a decrease in aggression; a positive coefficient indicates an increase in aggression.

Interpretation: There is no statistically significant treatment effect for aggression. The significant individual predictors are sex and fall baseline score. Males report one-fifth of a standard deviation more aggressive behaviors than females. For every standard deviation increase in the fall, students report about one-fourth of a standard deviation increase on aggression in the spring. This model accounts for two-thirds of the between school variance on aggression. It is possible that other school-level predictors would account for the remaining 1% of between school variance.

Table F8

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

Fixed effects

Variable	Coefficient[a]	SE	df	p
Intercept	-.09	.07	10	.21
Treatment status (0 = control, 1 = treatment)	.05	.13	10	.71
Individual sex (0 = female, 1 = male)	.24	.12	266	.04[*]
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.29	.18	266	.10
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.17	.10	266	.10
Individual fall rebellion score	.18	.06	266	.00[*]
MSA composite	-.19	.07	266	.01[*]

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.89	--	--	--
Level-2 intercept	.13	10	13.44	.20

Proportion of variance explained

Proportion of tau explained	.35
Proportion of sigma-squared explained	.18
Net ICC	.02
Proportion of between school variance explained	.33

Note. *SE* = standard error.
[a]A negative coefficient indicates a decrease in rebellious behavior; a positive
coefficient indicates an increase in rebellious behavior.

Interpretation: There is no statistically significant treatment effect for aggression. The
significant individual predictors are sex and fall baseline score. Males report one-
fourth of a standard deviation more rebellious behaviors than females. For every
standard deviation increase in the fall students report almost one-third of a standard
deviation increase on rebellion in the spring. This model accounts for one-third of the
between school variance on rebellious behavior. It is possible that other school-level
predictors would account for the remaining 2% of between school variance.

Table F9

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

Fixed effects

Factor	Coefficient[a]	SE	df	p
Intercept	-.04	.07	10	.63
Treatment status (0 = control, 1 = treatment)	.03	.15	10	.83
Individual sex (0 = female, 1 = male)	.48	.10	341	.00[*]
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.29	.18	341	.10
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.23	.11	341	.03[*]
Individual fall victimization score	-.00	.03	341	.97
MSA composite	.02	.06	341	.76

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.88	--	--	--
Level-2 intercept	.02	10	9.98	>.50

Proportion of variance explained

Proportion of tau explained	.01
Proportion of sigma-squared explained	.12
Net ICC	.03
Proportion of between school variance explained	.00

Note. *SE* = standard error.
[a]A negative coefficient indicates a decrease in victimization; a positive coefficient indicates an increase in victimization.

Interpretation: There is no treatment effect for aggression. The significant individual predictors are sex and FARMS status. Males report being victimized about half of a standard deviation more often than females. Students who receive reduced lunch report being victimized one-fourth of a standard deviation more often than those without free or reduced lunch. Students who receive free lunch report being victimized one-fourth of a standard deviation more frequently than students receiving reduced lunch (and half of a standard deviation more than those who do not receive either). By entering in individual predictors, all of the between school variance was explained. This model accounts does not account for any of the between school variance on victimization.

Table F10

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

Fixed effects

Variable	Coefficient [a]	SE	df	P
Intercept	.04	.06	10	.50
Treatment status (0 = control, 1 = treatment)	-.04	.11	10	.74
Individual sex (0 = female, 1 = male)	.37	.12	262	.00[*]
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian)	-.19	.19	262	.34
FARMS status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch)	.11	.11	262	.33
Individual fall acceptability of aggression score	.31	.06	262	.00[*]
MSA composite	.00	.07	262	.98

Random effects

Variance component	Variance	df	χ^2	p
Level-1 error	.91	--	--	--
Level-2 intercept	.02	10	9.59	>.50

Proportion of variance explained

Proportion of tau explained	.98
Proportion of sigma-squared explained	.15
Net ICC	.00
Proportion of between school variance explained	1.00

Note. *SE* = standard error.
[a]A negative coefficient indicates a decrease in acceptability of aggression; a positive
coefficient indicates an increase in acceptability of aggression.

Interpretation: There is no statistically significant treatment effect for acceptability of
aggression. The significant individual predictors are sex and fall baseline score.
Males report aggression as one-third of a standard deviation more acceptable than
females. For every standard deviation increase in the fall students report almost one-
third of a standard deviation increase on acceptability of aggression in the spring.

References

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.

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.

Conduct Problems Prevention Research Group (1999a). Initial impact of the Fast Track prevention trial for conduct problems: I. The high risk sample. *Journal of Consulting and Clinical Psychology, 67,* 631-647.

Conduct Problems Prevention Research Group (1999b). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648-657.

Conduct Problems Prevention Research Group (2000). Merging universal and indicated prevention programs: The Fast Track model. *Addictive Behaviors, 25,* 913-827.

Conduct Problems Prevention Research (2002a). Predictor variables associated with positive Fast Track outcomes at the end of third grade. *Journal of Abnormal Psychology, 30,* 37-52.

Conduct Problems Prevention Research (2002b). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Psychology, 30,* 1-17

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. (1997). Reinterpreting dissemination of prevention programs as widespread implementation with effectiveness and fidelity. In R. P. Weissberg, T. P. Gullotta, R. L. Hampton, B. A. Ryan, & G. R. Adams (Eds.), *Establishing preventive services* (pp. 253- 289). Thousand Oaks, CA: Sage.

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.

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.

Embry, D.D., Fkannery, D.J., Vazsonyi, A.T., Powell, K.E., & Atha, H. (1996). Peace Builders: A theoretically driven, school-based model for early violence prevention. *American Journal of Preventive Medicine, 12,* 91-100.

Flannery, D.J, Liau, A.K, Powell, K.E., Vazsonyi, A.T., Guo, S., Atha, H., & Embry, D. (2003). Initial behavior outcomes for the Peace Builders universal school-based violence prevention program. *Developmental Psychology, 39,* 292-308.

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.

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., & Skroban, S. (1998).Can prevention work where it is needed most? *Evaluation Review, 22,* 315-340.

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) (available on line at http://www.gottfredson.com)

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., Frey, K. , & Rivara, F.P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association, 277,* 1605-1611.

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.

Hartshorne, H.; May, M.A; & Shuttleworth, F.K. (1930). Studies in the organization of character. New York: The MacMillan Company.

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.

Institute for Education Sciences (IES, 2003). *Request for proposals: Social and character development research grants.* Released on January 15, 2003 on http://www.ed.gov/offices/IES/funding.html.

Jakob, J (2005). An evaluation of Second Step: A violence prevention curriculum with kindergarten students. *Dissertation Abstracts International, 66* (5), 638A.

Lai, E.L.Y. (2001). *Ethnographic evaluation of Second Step: A violence-prevention curriculum.* Unpublished doctoral dissertation, The University of British Columbia, Canada.

Linares, L.O, Rosbruch, N, Stern, M.B., Edwards, M.E., Walker, G., Abikoff, H.B., & Alvir, J.M.J. (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.

Martin, G. & Pear, J. (1999). *Behavior modification: What it is and how to do it* (6th Ed.). Upper Saddle River, NJ: Prentice Hall.

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.

Mrazek, P.J. & Haggerty, R.J. (Eds.) (1994). Reducing risks for mental disorders: Frontiers for preventive intervention research. Washington, DC: National Academy Press.

Newcomb, M.D., Abbott, R.D., Catalano, R.F., Hawkins, J.D., Battin-Pearson, S., & Hill, K. (2002). Mediational 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.

Nicolet, I.A. (2004). The Second Step violence prevention program: Effectiveness of a brief social skills curriculum with elementary-age children. *Dissertation Abstracts International, 65* (7), 3693B.

Oppitz, J.L. (2002). Violence prevention: Empowering school counselors. A study of strategies used by practicing elementary school counselors. *Dissertation Abstracts International, 63*(8), 2800A.

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. (2002). *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* (2nd ed.). Thousand Oaks, CA: Sage.

Reed, J.G. (2004). An examination of treatment integrity practices and behavioral outcomes when utilizing the Second Step curriculum. *Dissertation Abstracts International, 65*(3), 831A.

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.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rossi, P.H.; Lipsey, M.W.; & Freeman, H.E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Rubin, D.B. (1991). EM and beyond. *Psychometrika, 56,* 214-254.

Schochet, P.Z. (2005). *Statistical power for random assignment evaluations of education programs* (MPR Reference No. 6046-310). Princeton, NJ: Mathematica Policy Research, Inc.

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.

Spivack, G., Platt, J.J; & Shure, M.B. (1976). The problem-solving approach to adjustment. San Francisco, CA: Jossey-Bass Publishers.

Stern, M.B. (1999). Unique Minds School Program. Unpublished manual, Ackerman Institute for the Family, 149 East 78th Street, New York, NY 10021.

Teagarden, J.M. (2002). *Effects of the instructional delivery upon violence prevention in elementary-age students.* Unpublished doctoral dissertation, Kansas State University.

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.

Washburn, J.J. (2001). Evaluation of a violence prevention program with low-income, urban African American youth. *Dissertation Abstracts International, 62*(9), 4242B.

Weissberg, R.P; Kuster, C.B; & Gullotta, T.P. (1997). Introduction and overview: Prevention services- From optimistic promise to widespread, effective use. In R.P Weissberg, T.P. Gullotta, R.L Hampton, B.A. Ryan & G.R. Adams (Eds). *Establishing Preventive Services* (p.1-26). Thousand Oaks, CA: Sage Publications.

Wilke, K.L. (2004). *Exploring potential relationships among student, school, and counselor variables and violence prevention program outcomes in a district-wide elementary comprehensive school counseling program.* Unpublished doctoral dissertation, Seattle Pacific University.

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., & 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.

Wojtalewicz, M.P. (2004). Examination of Head Start children's social competence and social cognitions after participating in a universal violence prevention program. *Dissertation Abstracts International, 65* (12), 4465A.