ABSTRACT

Title of Document: AN EXPERIMENTAL EVALUATION OF THE EFFECTS OF A SCHOOL-BASED, UNIVERSAL PREVENTION PROGRAM ON PARENT AND TEACHER RATINGS OF STUDENT BEHAVIOR

Allison Joan Nebbergall, Doctor of Philosophy, 2009

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

Meta-analyses of skills-based prevention interventions show moderate effect sizes for increasing social competencies and decreasing behavior problems. While the literature suggests that prevention programs can be effective, rigorous independent research is lacking regarding the efficacy of many specific programs. The present study is based on a randomized-control experiment evaluating Second Step in 12 Maryland public elementary schools to assess the effects of the program on parent and teacher ratings of student behavior. Ratings using the Social Competency Rating Form had previously been considered as a single global measure of student behavior, and had not been found to be affected by the intervention. Nonetheless, a re-consideration of the psychometric properties of the scale and its sensitivity to skills taught by the Second Step curriculum led to the speculation that separation of the global measure to reflect distinct ratings of social competency and problem behavior might reveal effects on the social competency component. Analyses show no effects on parent or teacher ratings of social competency

or on teacher ratings of problem behavior. In some analyses, students in treatment schools had nearly twice the odds of being classified in a “problem” group according to ratings made by their parents than did students in control schools. Results were supported by sensitivity analyses using weights and imputation.

AN EXPERIMENTAL EVALUATION OF THE EFFECTS OF A SCHOOL-BASED, UNIVERSAL PREVENTION PROGRAM ON PARENT AND TEACHER RATINGS OF STUDENT BEHAVIOR

By

Allison Joan Nebbergall

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Gary Gottfredson, Chair
Associate Professor Cheryl Holcomb-McCoy
Professor Dennis Kivlighan
Professor Peter Leone
Associate Professor William Strein

Acknowledgements

This research is supported in part by grant number R305L030002A from the Institute of Education Sciences, U.S. Department of Education. Opinions expressed in this manuscript do not necessarily reflect the opinions or policies of the U.S. Department of Education. I would like to thank the members of the University of Maryland research team (Gary Gottfredson, Amy Silverman, Elise Harak, Joe Nese, and Fortune Shaw) who played an integral part in this research. I would also like to recognize the school counselors (Sharon Atkinson, Kristin Clute, Cindi Cramer, Rachel Dressin, Beth Edelstein, Jenn Elsis, Holly Kleiderlein, Patricia Lynch, Joyce O'Connell, Carole Peacock, Connie Poussard, Aileen Selznick, Dana Smith, Frances Walker, and Teresa Wismer) and support staff (Lucia Martin, Debbie Wooleyhand, and Ellen Wilson) whose hard work made possible the implementation of this research in the schools.

Correspondence concerning this manuscript may be addressed to Allison Joan Nebbergall, 3214 Benjamin Building, University of Maryland- College Park, College Park, MD 20742. Email: ajn@umd.edu.

Table of Contents

List of Tables

Literature Review

A first step in preventing problem behavior and delinquency is understanding the factors associated with the occurrence of these outcomes. Risk factors operating within multiple overlapping contexts contribute to one's likelihood of developing psychopathology, problem behavior, or substance abuse. Individual characteristics associated with problem behavior include early and persistent behavior problems (Spivack, Marcus, & Swift, 1986), hyperactivity (Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikstrom, 2002), rebellious behaviors and attitudes (Herrenkohl et al., 2000), and alienation from peers (Jessor et al., 1995). An individual's family and peer interactions can also put him or her at risk for future conduct problems. Risk factors associated with family environments include poor and unstable family management practices, family discord and conflict, parental violence and criminality, and residential mobility (Herrenkohl et al., 2000). Risk factors associated with peer interactions include early peer rejection (Consortium on the School-based Promotion of Social Competence, 1994) and associating with delinquent peers (Jessor et al., 1995; Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikstrom, 2002). School and community factors further influence a child's risk level. Risk factors related to school experiences include academic failure, poor school attachment, low achievement expectations (Herrenkohl et al., 2000), and low school motivation (Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikstrom, 2002). Community risk factors include poverty, disorganization, normative acceptability of problem behavior, and easy access to drugs and alcohol (Consortium on the School-based Promotion of Social Competence, 1994; Herrenkohl et al., 2000). As the number of

risk factors accumulates, independent of any particular risk factor, the chances of developing a disorder or engaging in delinquent behavior increases as well (Rutter, 1979).

Harmful effects of risk are thought to be mitigated by the presence of protective factors. Protective characteristics or conditions may help to explain how and why youths who have been exposed to the same risk factors may be differently affected (Rutter, 1987). Individual characteristics such as a resilient temperament, a prosocial orientation (Jessor et al, 1995), the capability to accurately interpret social cues, problem solving and decision-making skills, the ability to behave in a way that leads to desired outcomes, and a sense of self-efficacy (Linares et al., 2005; Scales, et al., 2000) may protect against risk. Positive family dynamics including supportive relationships, healthy bonding with adults, and low family conflict (Jessor et al., 1995) may also serve as protective factors. Finally, an environment that encourages and supports coping efforts and recognizes and rewards good behavior (Reiss & Price, 1996) may also buffer against risk. Like risk factors, protective factors have a cumulative effect. The more positive developmental assets that a person has, the less likely he or she is to develop psychopathology or engage in delinquent or aggressive behaviors (Leffert et al., 1998; Scales et al., 2000).

The parallels between risk and protection lead to disagreement about whether or not protective factors uniquely contribute to the study of risk. When comparing the list of risk and protective factors, protective factors appear to be antonyms for risk factors, suggesting that risk and protection may be opposite ends of a single continuum. For example, family conflict is a risk factor for future delinquency, whereas the absence of family conflict is a protective factor. Risk and protection may merely be a matter of

semantics. From this perspective, it would be more parsimonious to consider risk and protection as a single construct. On the other hand, this may be an oversimplified argument when one moves beyond solely identifying risk and protective factors to also consider the processes through which risk and protection operate (Rutter, 1987). From this perspective, protective factors represent more than the absence of risk. Protective factors operate in the context of risk; they may mediate or moderate the relation between risk factors and behavior (Pellegrini, 1990).

Jessor et al. (1995) have tried to test these claims, and assert that protective factors can exert their own distinct direct effects on behavior, as well as moderate the relation between risk factors and behavior. Using a measure of Friends as Models for Problem Behavior (a risk factor) and Friends as Models for Conventional Behavior (a protective factor), they found that these variables were only modestly correlated (-0.20) and that both contributed unique variance to the problem behavior outcome in a final regression model. A subsequent study found a similar relation. Jessor, Turbin, and Costa (1998) found that separate measures of risk and protection accounted for relatively comparable and significant contributions in accounting for negative and successful outcomes among disadvantaged youth. However, this evidence should be interpreted cautiously, as the issue of opposite wording still applies. That is, if the protective factor used were simply reflected, the incremental R^2 reported would then reflect the contribution of two risk factors rather than one risk and one protective factor.

More compelling would be evidence of an interactive relationship between risk and protection. Jessor et al. (1995, 1998) conclude that protection significantly moderated the effects of risk in their studies. In these two studies, the R^2 change ranged from 0.004

to 0.02, small values that may be elevated as a result of modeling possibly nonlinear bivariate relationships. Therefore, the R^2 change values may be too small to confidently claim a clear demonstration of a meaningful interaction of risk and protection. The relation between risk and protection still requires additional empirical exploration.

Many prevention programs include a focus on both curtailing behavioral dysfunction and promoting social and emotional development (Durlak & Wells, 1997). Among the most widely used and recommended types of interventions are those focused on teaching children social competencies such as problem-solving, decision making, social approach and engagement, and communication skills (Consortium on the School-based Promotion of Social Competence, 1994). For example, children with conduct problems may benefit from training aimed at the development of positive prosocial skills, such as problem-solving, anger management, appropriate play interaction, and empathy (Webster-Stratton & Lindsay, 1999). Social competencies help youths integrate thoughts, feelings, and actions to achieve desired social and interpersonal goals. Pertinent skills include identifying and accurately interpreting relevant social cues, effectively handling interpersonal conflicts, realistically anticipating consequences and potential obstacles to goal attainment, and understanding the link between desired outcomes and actions necessary to achieve those outcomes (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002). These types of competencies have been suggested as protective factors for resilient youth. The direct promotion of these social competencies has been shown to be useful in the prevention of childhood psychosocial problems such as delinquency and drug use (Chung & Elias, 1996), the promotion of academic adjustment (Gresham & Elliott, 1990), and in the treatment of behavioral and emotional problems such as

oppositional defiant disorder, conduct disorder, depression, and anxiety (Vera & Gaubatz, 2002).

School-Based Prevention Programs

Schools are a central setting in which prevention programs can be implemented. Schools have consistent access to youth throughout their development, and also offer a pipeline through which most parents and guardians can be contacted. Schools are a venue in which students can develop and practice self-control and build socialization skills, factors that may help to lessen the risk of later delinquency. It is important to begin teaching these skills in elementary school before students are pressured to engage in risky behaviors and to continue teaching them through high school (Gottfredson, 2001).

Schools acknowledge the need for prevention efforts and recognize their potential in positively affecting youth outcomes. As such, school-based prevention efforts are pervasive. Eighty percent of a national sample of elementary schools reported using a curricular prevention program in response to behavior problems, with a median number of 14 different prevention activities occurring concurrently (G. Gottfredson, D. Gottfredson, Czeh, Cantor, Crosse, & Hantman, 2000). Unfortunately, despite good intentions, these prevention programs are often characterized by multiple fragmented initiatives not in concordance with the school's mission, culture, or needs, and by poor staff training and inconsistent, inadequate implementation (Greenberg et al., 2003). Even research-based initiatives may not be transferable to all real-world school settings. The neediest schools and school districts may be insufficiently equipped to implement program components with the quality and intensity necessary to achieve desired outcomes (D. Gottfredson, G. Gottfredson, & Skroban, 1998). Programs that are

insufficiently organized, monitored, evaluated, and improved over time have less of an influence on student behavior and are unlikely to be maintained. These uncoordinated efforts may not only be ineffective, but also have the potential to be disruptive (Greenberg et al., 2003).

Nation et al. (2003) conducted a review-of-reviews across four areas (substance abuse, risky sexual behavior, school failure, and juvenile delinquency and violence) to identify and synthesize the characteristics of effective prevention programs. Nine principles associated with effective prevention programs were identified. These principles were related to three broad areas of prevention programming: program characteristics, matching programs to the target population, and implementing and evaluating prevention programs. A comprehensive curriculum was the program characteristic associated most strongly with effective prevention programs. Successful comprehensive programs include a combination of interventions (e.g., promotion of awareness and knowledge, development of skills) across the multiple settings that may curb the development of targeted problem behavior. Successful programs are also multi-faceted, incorporating a variety of methods, including an active skills-based component and hands-on practice or activities. The program should also be of sufficient length, which includes consideration of the duration, frequency, and spacing of the intervention. There should also be research-based justification of a prevention program to expect that it will produce the desired changes in the targeted behavior. Providing the opportunity for youth to develop positive relationships was also associated with effective programming. When considering program match, the area most strongly associated with effective programming was that the intervention was appropriately timed. That is, programs must target developmentally

appropriate outcomes and have developmentally appropriate curricula. Also, certain types of programs may be differentially effective for different populations or different outcomes, so there must be a match between the program type, the population it targets, and the desired outcome. Certain characteristics related to implementation and evaluation were also identified. A successful intervention includes formal evaluations of its processes and outcomes. Also, staff who implement the program must be well trained, which may include follow-up trainings in addition to initial training (Nation et al., 2003).

Prevention efforts are not only intended for at-risk youth; they may be targeted at populations with varying levels of risk. Primary (also called universal) prevention refers to interventions with an entire population that has not already manifested the targeted problem. Secondary (also called selective or targeted) prevention refers to interventions with subgroups of the populations that have risk factors for a particular problem or that have already displayed minor problems. Finally, tertiary (also called indicated or individual) prevention refers to interventions with people who are at high risk or who have already demonstrated the identified problems (Mrazek & Haggerty, 1994; Strein & Koehler, 2008).

Primary interventions are delivered to all students in a school or grade. All students can benefit from the development of prosocial skills, which improve their overall psychological and social well-being and thus better prepare them to withstand or cope with potential risk factors for maladjustment. Universal prevention programs often seek to reduce risk for later problem behavior by simultaneously emphasizing the reduction of problem behaviors and the enhancement of social and emotional competencies (Durlak & Wells, 1997). This dual focus on reducing problem behavior and

developing social competencies may be more effective in promoting positive youth development and reducing problem behaviors than focusing only on strengthening competencies (Catalano, Hawkins, Berglund, Pollard, & Arthur, 2002).

Meta-analyses of skills-based prevention interventions have shown moderate effect sizes for increasing social competencies and decreasing behavior problems (Consortium on the School-based Promotion of Social Competence, 1994). Durlak and Wells (1997) conducted a meta-analysis of 177 primary prevention programs, 73% of which were delivered in schools, designed to prevent problem behaviors and increase social competencies. The mean effect size across studies was 0.34, with a range of -0.45 to 2.36, with 9 of the 177 programs having a negative effect size. Most types of primary prevention programs achieved statistically significant positive effects in reducing negative behaviors, such as anxiety, depression, and problem behaviors, and in increasing competencies, such as assertiveness, communication, and self-confidence. Most types of programs were not differentially effective for problem behavior and social competency outcomes.

Wilson and Lipsey's (2007) most recent meta-analysis of school-based violence prevention programs examined program effects on outcomes such as aggression and violence, social skills, academic performance, and self-esteem. A total of 399 studies using an experimental or quasi-experimental design were analyzed. Across all outcomes, there were positive and statistically significant average effect sizes ranging from 0.20 to 0.35. Specifically for aggressive/disruptive behavior and social skills, the mean effect sizes were 0.21 and 0.32, respectively. The effect size on aggressive/disruptive behavior was the same, 0.21, for the subset of 77 universal prevention programs.

An evaluation of 28 studies in elementary schools provides further support for the effectiveness of universal school-based programs in reducing violence-related behaviors and increasing social competencies (Hahn et al., 2007). The interventions included were predominantly classroom-based and teacher-delivered, focusing on reducing disruptive and antisocial behavior and on developing social skills and promoting environmental change. Effects were characterized as a percentage change in the occurrence of violent behavior in the intervention group relative to the control group. In elementary school programs, the median percentage change was an 18.0% (interquartile range of -44.8% to -2.5%) reduction in violent behavior. Results were similar for different ethnic and socioeconomic samples and for programs using different intervention strategies (e.g., informational, cognitive/affective, and social skills building).

Each of these meta-analyses used different selection criteria and analytic strategies. However, despite these methodological differences, the outcomes are consistent overall. These studies concluded that there is evidence in support of the effectiveness of school-based violence prevention programs in reducing problem behaviors and increasing social competencies. The studies all also point to some gaps in the literature. Future research should focus on improving efforts to outline intervention goals and procedures, assess the amount and quality of implementation, determine effectiveness of real-world practices outside of research demonstrations, identify if programs are differentially effective for different outcomes or populations, and conduct long-term follow-up to determine whether program benefits are maintained (Durlak & Wells, 1997; Wilson & Lipsey, 2007).

Second Step: A Violence-Prevention Curriculum

Although the literature suggests that prevention programs can, in general, be effective, more rigorous independent research should be conducted regarding which specific programs are effective. The present study evaluates the effect of one school-based universal prevention program in particular, Second Step: A Violence Prevention Curriculum (Committee for Children, 2002), after three years of implementation in six elementary schools by focusing on teacher and parent reports of problem behavior and social skills.

Second Step is one of three programs published by the Committee for Children, whose programs are used in 25,000 schools in 21 nations and regions around the world (Committee for Children, 2009). The Second Step curriculum is widely popular, with a market saturation of approximately 35% overall in the United States (B. Bailey-Lewis, personal communication, May 8, 2009). Second Step has the dual goals of reducing aggressive and disruptive behaviors and fostering the growth of social competencies (Committee for Children, 2002). The curriculum is primarily based on social learning theory, which emphasizes learning in a social context, including modeling, role-play and practice, and observational learning. The curriculum also draws from social information-processing, cognitive-behavioral, and self regulatory models. Concepts and strategies from these frameworks are integrated into the curriculum's three core content areas: empathy, problem-solving and impulse control, and anger management (Frey, Hirschstein, & Guzzo, 2000).

The Second Step lessons are delivered in the classroom by teachers or school counselors, with a lesson generally occurring weekly or semiweekly throughout the

school year. The elementary curriculum spans kindergarten to fifth grade. The core content areas are the same for each grade, but the specific skills and language are particular to each grade level. At each grade level, there are 15 to 22 lessons lasting approximately 30 minutes per lesson. Every lesson begins with and is structured around a scripted photo-lesson card accompanied by a social vignette. In addition to a suggested script, each card contains key concepts, lesson objectives, and activities for the program implementer. Teachers read the scenario accompanying the photo-lesson cards and guide whole group discussion and activities. Videos and posters supplement many of the lessons. Each lesson is characterized by similar key strategies including whole-group discussion, modeling, role-play, and transfer of learning strategies intended to promote generalization of skills (Frey, Hirschstein, & Guzzo, 2000).

Second Step has been the subject of several evaluations. The studies reviewed here were chosen because they were published in peer-reviewed journals, took place in elementary grades, and examined outcomes similar to those of interest in the present study. Grossman et al. (1997) conducted the only fully randomized control trial. The study was conducted in 12 schools in 49 second and third grade classrooms with 790 consented students. Following one year of implementation, there was little change in intervention and control schools on parent and teacher reports of student social competence and aggressive behavior. However, there was a significant decrease in observed physically aggressive behavior and a significant increase in neutral/prosocial behavior among children receiving the curriculum compared with children in the control group. These differences were observed on the playground and in the cafeteria, but not in the classroom.

While important as the only randomized control-group study, the Grossman et al. study still has several notable limitations. First, the program was implemented in only approximately four classrooms in each of the schools. Second Step is a whole-school program, so isolated implementation in only selected classrooms is not reflective of a comprehensive implementation of the Second Step curriculum. Also, the intervention was implemented for only one year, and it is possible that a multi-year implementation would be necessary to produce effects. No assessments of implementation were conducted, and it is therefore unknown if the minimal effects can be attributed to low program implementation. Finally, with only 12 schools in the study, the power to detect effects is relatively low.

Frey et al. (2005) conducted a partially randomized evaluation of the Second Step program over two years. In the first year, 11 of the schools were randomly assigned to the intervention (eight schools) or control group (three schools). The following year four additional schools were recruited and assigned to the control group, for a total of 15 schools. This method of assignment precludes the study from being considered a true experiment. Study participation was limited to second and fourth grade classrooms. However, in both the treatment and control schools, the classrooms not participating in the study received Second Step curriculum materials, teacher training, and substitute teachers. Potential treatment contamination across classrooms in the control condition is possible, especially since Second Step is intended to be a whole-school intervention.

Results from this study were mixed. After one year of implementation, teacher ratings of student antisocial behavior decreased more for the intervention group than for the control group, and teacher ratings of student social competence increased for the

intervention group relative to the control group. After two years of implementation, there were no treatment and control group differences in teacher ratings of antisocial behavior. Teacher ratings of student social competence were still higher for the intervention group relative to the control group, but to a lesser degree than the differences observed in the first year. Student surveys of hostile attributions and intentions, as measured using vignettes of ambiguous and non-ambiguous situations, showed no significant group differences.

The inconsistent effects do not provide clear evidence that Second Step is effective at reducing problem behavior or increasing social competence. Preliminary positive results diminished over two years of implementation rather than increased. Furthermore, while the study had a large sample size, there was a high percentage of non-consenters (37%) and high attrition over the two years (25.5% in the intervention group and 28.8% in the control group). Also, while no formal measurement of implementation was conducted, the authors note that implementation was incomplete, with a minority of students receiving the total program.

Holsen and Frey (2008) used a quasi-experimental age-cohort design to study the effect of the Second Step curriculum on student reports of social competencies and problem behaviors when delivered to fifth and sixth grade students in Norwegian elementary schools. The sample consisted of 11 schools with 1153 students participating at baseline. The authors reported varied implementation among schools, modification of lessons by teachers, and varying levels of administrative support. However, there were no measures of implementation to explore the degree or influence of weak implementation.

The outcome data were analyzed in two different ways, using a repeated-measures analysis and an age-cohort design. Both methods revealed some scattered positive results for certain grades or sexes. Specifically, the repeated measures analysis showed that fifth grade students increased their self-reported social competence after the first year of program implementation, whereas the sixth grade students did not. There were no effects on self-reported internalizing and externalizing problem behaviors. The analysis of the data using age-cohort methodology yielded similar results. The post-test level of social competency was higher in the sixth grade intervention group than in the comparison group. There was an increase in self-reported social competence in the seventh grade intervention group relative to the comparison group, but only for girls. There was a significant decrease in externalizing problem behavior for boys in the sixth grade intervention group relative to their comparisons. There were no differences for self-reported internalizing problems. The different results by outcome, grade, and sex, coupled with the weak implementation and non-experimental design, make it difficult to draw any conclusions from this study regarding the effectiveness of the Second Step program.

Cooke et al. (2007) used a pretest-posttest design to study the influence of Second Step on third and fourth grade students' prosocial and aggressive behaviors. The evaluation was conducted in five schools with a total of 741 participants. For this study, the Second Step curriculum was modified to include a larger community component, thereby creating a city-wide implementation. However, there were no measures of implementation fidelity.

Assessments were conducted immediately prior to and following implementation of the curriculum using student self-report questionnaires, behavioral observations using a checklist procedure adopted from Grossman and colleagues (1997), and a review of disciplinary referrals. Mixed positive and negative effects were found on the student self-report questionnaire. There were significant increases in positive approach/coping, caring/cooperative behavior, suppression of aggression, and consideration of others, no changes in responsibility, and a significant decrease in impulse control. There was a small increase in self-reports of angry and aggressive behaviors. Behavioral observations failed to replicate the findings reported by Grossman et al. (1997); there were no changes in violent or aggressive behavior or in neutral behaviors, and a small decrease in positive, borderline, and negative behaviors. No change in discipline referrals was found. Due to the mixed positive and negative effects, weak pre-test post-test design and small number of schools, this study does not contribute clear evidence that Second Step improved or worsened student behavior.

McMahon and Washburn (2003) used a pretest-posttest design to examine the effectiveness of Second Step with low-income, urban African American youth in increasing knowledge about violence, increasing prosocial behavior, and decreasing aggressive behavior. Second Step was delivered for one year to students in grades five through eight in two schools, with a total of 156 students at baseline. Outcomes were assessed using student self-reports of aggression, empathy, impulsivity, and sense of school membership, and teacher and peer ratings of aggression and prosocial behavior. A repeated-measures ANOVA found that students gained knowledge concerning violence, the consequences of violence, and the prevention of violence, but there was mixed

evidence that students demonstrated behavior changes. There were no consistent effects on teacher ratings of aggression or prosocial behavior; a decrease in ratings was found in one school and an increase in ratings was found in the other school. There was an increase in student self-reports of empathy, but no changes on reports of aggression, impulsivity, and sense of school membership. Again, this study used a weak pretest-posttest design and found both positive and negative results for student behavioral outcomes. This study does not provide evidence that Second Step has a positive or negative effect on student behavioral outcomes.

Taub (2001) studied the effects of Second Step in a rural, white, low-income elementary school. In one school, one classroom in grades three, four, and five was randomly selected to participate in the study, resulting in a total of 54 students in the treatment condition. In a comparison school, 33 students from two classrooms were selected. Classroom observations and teacher ratings of student social competency and problem behavior were collected prior to the start of the intervention, prior to the end of the school year, and one year after implementation of the intervention. The two samples were found to be nonequivalent at pretest, with the students in the treatment schools being rated as demonstrating fewer socially competent behaviors and more antisocial behaviors than the children in the comparison school. Following one year of implementation, there was an increase in teacher ratings of social competence in the intervention group, resulting in treatment and control students being rated similarly by their teachers after a year of the Second Step program. Teacher ratings of antisocial behavior increased from Time 1 to Time 2 for both treatment and control students. By Time 3, treatment students were rated as demonstrating fewer antisocial behaviors than

they had at pretest, whereas control students were rated as demonstrating more antisocial behaviors than they had at pretest. Behavior observations showed inconsistent results over time and across behaviors. Inconsistent changes between Time 1, Time 2, and Time 3 also make it difficult to interpret changes in the measured outcomes. This study is further limited by the small sample size and lack of random assignment of participants to the treatment condition. The lack of random assignment, compounded by the large treatment-control differences at pretest, makes it impossible to draw casual inferences regarding the Second Step program.

McMahon, Washburn, Felix, Yakin, and Childrey (2000) used a pretest-posttest design to examine the effects of Second Step on 109 urban preschool and kindergarten students. Effects were assessed using child interviews, teacher ratings of social skills and problem behaviors, and behavioral observations of disruptive behavior, verbal aggression, and physical aggression conducted at the classroom-level. Results showed that the students gained knowledge in identifying feelings and facial cues, in thinking about how children might respond in conflict situations and why, and in predicting the consequences of their responses. Inconsistent results were found for teacher ratings of student behavior. Teacher ratings of problem behaviors decreased for preschool students and increased for kindergarten students, while teacher ratings of social skills did not differ significantly over time. Behavior observations revealed decreases in all three negative behaviors, with the decrease in disruptive behavior being more pronounced for kindergarten students than for preschool students. However, observations were conducted at a classroom-level in only four classrooms, and do not reflect changes in individual student behavior. This study was limited by the small, transient sample and lack of a

control group. Without a control group, it is impossible to know if any of the demonstrated changes were caused by the Second Step program.

The studies reviewed show that the literature evaluating the efficacy of the Second Step program largely suffers from weak non-experimental designs, inconsistent results for different behavioral outcomes and sample subgroups, and failure to measure implementation fidelity. These studies do not meaningfully provide support for or against the effectiveness of the Second Step intervention.

Present Study

The purpose of the present study is to evaluate the effects of Second Step on parent and teacher ratings of student social competency and problem behaviors. This study builds upon the literature by filling in several gaps. First, it is a randomized control group trial conducted by an independent group of researchers, which allows for an independent estimate of effects of the program. Second, it is a multi-year evaluation studying the effects of the curriculum for students who received three years of whole-school implementation. Third, effects are measured using a scale closely aligned with the skills taught by the Second Step curriculum. Finally, formal implementation data are collected and show that the program was fully and adequately implemented (See Appendix A for a summary of program delivery).

This study expands upon the prior analyses conducted within this project (Gottfredson et al., 2008) by subdividing the global measure of social competencies and problem behaviors used earlier into separate social competency and problem behavior scales to be used as dependent variables. The Second Step lessons target specific social skill areas. Students are taught empathy skills including how to identify their emotions,

take the perspective of others, and respond in a caring way. They also learn to react thoughtfully rather than impulsively by using problem-solving steps and anger management techniques (Committee for Children, 2002). It is essential that the measure used to assess the effectiveness of a program is sensitive to the treatment effects of interest (Lipsey, 1983). The social competence items of the Social Competency Rating Form (SCRF) relate directly to the specific skills taught in the program. Isolating the social competence items as a separate scale may result in greater sensitivity to behavior change in the skill areas targeted by the curriculum. By examining parent and teacher ratings using the SCRF as a global social competence outcome, the previous analyses did not give the program the best chance of demonstrating potential effects. Also, problem behavior ratings are markedly skewed. When the ratings were considered as a single scale, the overall distribution was skewed. There was a potential ceiling effect caused by inclusion of the problem behavior items, as most students were rated as displaying few problem behaviors. When there is a ceiling or floor effect, a change on the construct may not be detected by the measure (Lipsey, 1983). Parceling out the problem behavior items reduces the possibility that a ceiling effect is masking potential treatment effects. Furthermore, the standard HLM analyses originally run are appropriate when there is a continuous outcome that is normally distributed (Raudenbush & Bryk, 2002). The normality of the social competence scale is improved when the problem behavior items are parceled out. Separating the two scales creates a normally distributed social competency distribution and a dichotomous problem behavior outcome, thereby creating a dataset that better conforms to the statistical assumptions required of the data analysis methods.

This study asks the following questions:

1. Does the Second Step program significantly increase or decrease teacher ratings of student problem behaviors among students in treatment schools compared to those in control schools?
2. Does the Second Step program significantly increase or decrease teacher ratings of student social competencies among students in treatment schools compared to those in control schools?
3. Does the Second Step program significantly increase or decrease parent ratings of student problem behaviors among students in treatment schools compared to those in control schools?
4. Does the Second Step program significantly increase or decrease parent ratings of student social competencies among students in treatment schools compared to those in control schools?
5. Are there any interactions of treatment with individual student demographic characteristics (i.e., ethnicity, FARM eligibility, grade level) and baseline scores for any of the outcome variables examined?

This final research question is included as a supplementary area of interest. Due to the small number of minority group students in the sample and the few degrees of freedom at level-2, the sample is inadequate for fully exploring the possibility of interaction effects.

Methods

This study used data collected as part of a larger project evaluating the effect of Second Step on multiple behavioral and academic outcomes. This study narrows the focus by specifically examining the effects of Second Step on parent and teacher ratings of student social competency and problem behavior.

School Selection and Random Assignment

The sample includes two cohorts of students who received three years of the Second Step curriculum in 12 public elementary schools in a Maryland school district. The school district includes mostly suburban communities, with a total of 109 schools and approximately 74,000 students. Of these 74,000 students, approximately one third are members of ethnic minority groups and approximately one fifth participate in the free and reduced meals program. The 12 schools in the study were selected for participation because they had not previously implemented the Second Step program and agreed to be randomly assigned to the treatment or control condition. The 12 schools were divided into six matched pairs. Pairings were made based on school size, ethnic/racial composition, and the percentage of students participating in the Free and Reduced Meals (FARM) program. Next, one school of each pair was randomly assigned to either the treatment or control condition. The schools in the treatment condition participated in the whole-school implementation of the Second Step program, delivering the instruction in all general education classrooms in grades 1 – 5. The schools in the control condition continued to implement the programs already in place prior to the research project.

Table A2 of Appendix A displays information about the types of character education activities and materials used in the treatment and control schools. Both

treatment and control schools were implementing character education programs and activities. However, there was a large difference in the type of programming being implemented. Instruction and materials in treatment schools were more formal and lessons were delivered more frequently. Also, treatment teachers implemented key program components such as modeling, role play, and feedback much more than teachers in control schools.

Participants

In the 2004 – 2005 school year, affirmative written parental consent for participation was sought from all 1st, 2nd, 4th, and 5th grade students (except a small number of students in self-contained special education classrooms). In Fall 2005 and Fall 2006, consent was again sought from students who had not yet replied and upon registration for students new to the school. Data obtained from third grade students in 2004 – 2005, fourth grade students in 2005 – 2006, and fifth grade students in 2006 - 2007 are used as part of a larger multi-site project. These students were given different measures, and therefore are not included in these analyses. Of the students present at baseline, only those students in the first and second grade cohorts were eligible to receive three full years of treatment. Older students graduated prior to completion of the three full years of treatment. Outcome analyses are based on the sample of students who were present at baseline and exposed to three full years of treatment.

Rating scales were prepared by the research team and then delivered by the school counselor to each classroom teacher. Teachers received a packet of teacher surveys for consented students in their classroom, and also distributed labeled parent surveys to each consented student. Completed teacher surveys were returned in a sealed envelope to the

school counselor. Completed parent surveys were mailed back to the university or returned in a sealed envelope to the classroom teacher.

The overall consent rates and return rates for parent and teacher ratings for the baseline and final data collection periods are displayed in Table 1. In Year 3, of the parent and teacher ratings completed, about 51% were ratings of male students and 49% were ratings of female students. Approximately 90% of completed surveys were ratings of White/Non-Hispanic students, 4% of African American, 3% of Hispanic, 2.5% of Asian/Pacific Islander, and 0.5% of American Indian/Alaskan Native students. Approximately 4% of completed surveys were ratings of students who qualified for reduced meals and 5% were ratings of students who qualified for free meals. These are the demographic characteristics of the sample on which outcome analyses were run. A comparison of student characteristics in the baseline and follow-up samples is included in Appendix B.

Table 1

Consent Rates and Return Rates

				Completed survey for percentage of	
Year and rater	Students enrolled	Students consented	Survey completed	Students enrolled	Students consented
Fall 2004					
Teacher	1810	1147	1097	61	96
Parent	1810	1147	1007	56	88
Spring 2007					
Teacher	1456	1073	1001	69	93
Parent	1456	1073	832	57	78

Two-level HLM models with the student at level-1 and the school at level-2 were run to describe how attrition affected the sample. A first set of HLM analyses was conducted to determine if there was an interaction of treatment with attrition from the study. For those first and second grade students who were enrolled in Fall 2004, a dichotomous dependent variable was created to indicate whether or not a useable teacher or parent survey was returned three years later in Spring 2007. Treatment assignment was entered at level-2, and no covariates were included at level-1. Treatment assignment was not significant in predicting the likelihood of receiving a teacher rating (odds ratio = 1.35, *CI* = 0.73 - 2.45, $p = 0.30$) or a parent rating (odds ratio = 1.32, *CI* = 0.78 - 2.24, $p = 0.27$) in Spring 2007, indicating that there is no evidence of differential attrition from the study.

Analyses were also run to determine whether treatment assignment interacted with student characteristics to predict attrition. In each of the analyses, the level-1 predictor (i.e., sex, ethnicity, and FARM status) was tested for random slope, and treatment was used to model that slope. A significant interaction would indicate that certain groups participated at different rates in treatment and control schools. However, no level-1 predictors had significant random slopes, indicating no evidence of differential attrition.

Individual student characteristics were examined to determine if they were predictive of returning a survey. Separate analyses were run for each demographic variable of interest—sex, ethnicity, and FARM status. Results showed no differences in participation based on student sex. That is, males and females were equally likely to participate in the study. However, there was a difference in participation based on student

FARMS status for the receipt of parent surveys but not teacher surveys. Students in the FARM program were less likely to have received affirmative consent and a completed parent survey. Additionally, for both parent and teacher surveys, there was a difference in study participation based on student ethnicity. Students from academically disadvantaged ethnic minority groups and students in the FARM program were less likely to have parent report data in Spring 2007. Students from academically disadvantaged ethnic minority groups were less likely to have teacher report data in Spring 2007. These sets of analyses collectively indicate that while attrition restricts generalizability of the results, it is not a threat to internal validity. In order to address these issues, sensitivity analyses were conducted using weighting and imputation.

Intervention

The Second Step curriculum was delivered in the six treatment schools for three years beginning in Fall 2004. Counselors and school staff were trained on program delivery prior to the beginning of the 2004 – 2005 school year, with teachers new to the school being offered training at the beginning of each subsequent school year.

The Second Step curriculum was delivered by teachers in all general education treatment-school classrooms. Schools were provided with a suggested pacing guide to keep them on track for full program completion prior to the end of the school year. On average, teachers delivered one lesson per week, spending approximately 30 minutes on each lesson. Contemporaneous implementation logs were completed by teachers following completion of each lesson. On these logs, teachers indicated whether or not they completed each of the lesson's key activities (e.g., modeling, role-play, etc.), also indicating if they completed the activity exactly as prescribed or with modifications.

Monthly meetings were held with the research staff and the school counselors to provide feedback on the extent and quality of implementation.

Character development teams were also formed in each of the treatment schools. These teams were headed by the school counselors and were intended to aid in the whole-school implementation of the Second Step objectives.

Variables

Parent and Teacher Ratings. Parent and teacher ratings using an adaptation of the Social Competency Rating Form (SCRF; G. Gottfredson et al., 2002; Nebbergall, 2007) were the outcome variables in this study. The SCRF consists of 29 items, with 17 positively worded items and 12 negatively worded items. Sample items include: *Hits, kicks at, or jumps on other children; If provoked by peers, shows self-control; Solves problems with peers through compromise or discussion;* and *Expresses concern for others.* All items are answered on a 4-point Likert-type scale, with a 1 indicating "Almost Never," 2 indicating "Sometimes," 3 indicating "Often," and 4 indicating "Very Often." The SCRF is intended to be used as a research tool in studying social competence and problem behaviors, especially in the specific context of evaluating intervention programs. In response to rater feedback, the response scale was changed slightly in Spring 2005; "Almost Never" was changed to "Never/Almost Never." The revised scale was used in all subsequent data collection periods.

On the basis of a factor analysis, the items of the rating form are divided into two scales: Social Competency and Problem Behavior. Each scale is scored by taking the average rating for the items completed. Seventeen items had substantial loadings on the first factor; all of these items relate to social competence. Eleven items had substantial

loadings on the second factor; all of these items relate to problem behavior. The item *Is teased, hit, or bullied by other kids* was not included in the analyses, as it directly reflected neither a social skill nor a problem behavior. All analyses were conducted for the two scores reflecting separate ratings of student social competence and problem behavior.

The psychometric properties of the SCRF were recently examined in a thesis (Nebbergall, 2007). At that time I decided that a single global index score rather than two distinct factors best described the data. In the initial exploration of the SCRF, an exploratory factor analysis using principal axis factoring was conducted, and the scree plot of eigenvalues was examined to determine the number of factors to extract. Factor matrices were rotated orthogonally using the varimax method. Nebbergall (2007) stated that a visual analysis of the scree plot of eigenvalues showed that the scree began at Factor 3, thereby indicating that two factors should be extracted. However, upon examination of the items, it was observed that one factor was composed of all positively worded items and the other factor was composed of all reverse-scored items. While the use of mixed item wording assumes that respondents will answer both item types as if the items represent the same construct, including a mixture of positively and negatively worded items may confound the interpretation of a rating scale's factor structure. Item wording may result in a reverse coding method factor. As such, the appearance of two factors may represent two distinct underlying constructs or, at least partially, an artifact of item wording (Magazine et al., 1996). For example, one study of a teacher self-efficacy rating scale tested this theory and found that when items were modified so that both positive and negative wording was available for each item, the scale went from two

factors to one factor (Deemer & Minke, 1999). Therefore, taking the more cautious and parsimonious perspective, I concluded in my thesis that "the two factors [of the SCRF] were methodological artifacts rather than two separate constructs" (Nebbergall, 2007, p. 24).

Table 2

Items Composing the Scales of the Social Competency Rating Form

Social Competency scale	Problem Behavior scale
• Articulates different ways to solve a problem	• Acts without thinking
• Asks adult for help or advice about ways to resolve difficult situations	• Acts in ways that annoy or bother others
• Expresses concern for others	• Gossips or spreads rumors
• Helps others	• Hits, kicks at, or jumps on other children
• If provoked by peers, shows self-control	• If upset, responds with verbal aggression (swearing, calling names)
• If angered, expresses anger without being aggressive or destructive	• Is impulsive in interacting with peers
• Is able to see things from other children's perspectives	• Responds with physical aggression to problems with peers
• Lets others know how he/she feels about situations	• Shows defiance in interactions with adults
• Removes him or herself from potential problem situations	• Takes or steals things that belong to others
• Resists peer pressure when appropriate	• Teases, insults, provokes, or threatens others
• Shows respect for others	• Tells lies or cheats
• Solves problems with peers through compromise or discussion	
• Takes time to calm down when dealing with problem situations	
• Takes other peoples feelings into account before acting	
• Takes responsibility for own actions (for example, apologizes)	
• Tries a new approach to a problem when first approach is not working	
• Understands the likely consequences of his or her own actions	

However, for the current study, the literature and original dataset were re-examined, and I decided that the decision to study two distinct factors is practically and psychometrically defensible. Table 2 displays the items that compose each factor and Appendix C displays the component matrices for parent and teacher ratings using the

two-factor solution. It is true that one factor contains the positively worded items and that the other contains the reverse-scored items, but, when examining the content of the factors, it also appears that one factor relates to social competency and the other factor relates to problem behavior.

Of the previous Second Step efficacy studies that were reviewed, every one that assessed effects on teacher or parent ratings did so by reporting separate ratings of social competency and problem behavior (Frey et al. 2005; Grossman et al., 1997; McMahon & Washburn, 2003; McMahon, Washburn, Felix, Yakin, & Childrey, 2000; Taub, 2001). Easier comparisons of this study's results to prior research can therefore be made by also separating ratings. Furthermore, in previous Second Step efficacy studies, different, although inconclusive, results were found for measures of social competency and problem behavior. This suggests that it is possible that Second Step may have differential effects on student social competency and problem behaviors. If there are different effects on social competence and problem behavior, separating the measure along these lines will allow for such effects to be detectable. Furthermore, the items composing the social competency scale directly reflect the skills taught in the Second Step curriculum. If the intervention is effective, this scale should be maximally sensitive to behavior changes targeted by the curriculum. Therefore, splitting the parent and teacher report measures into two parts gives the Second Step program a fair chance to demonstrate effects using a maximally sensitive measure.

Correlations between the scales were calculated using parent and teacher survey data collected in Fall 2004 to determine the amount of shared variance between the factor-based scales. For teacher ratings, the Social Competence scale and the Problem

Behavior scale have a correlation of -0.64. Corrected for attenuation due to unreliability (calculated using $r_{xy,corrected} = r_{xy} / ((r_{xx})(r_{yy}))^{½}$), the correlation is -0.69. For teacher ratings collected in Spring 2005, 2006, and 2007, the Social Competence and Problem Behavior scales have true-score correlations of -0.74, -0.76, and -0.74, respectively. For parent ratings, the Social Competence scale and the Problem Behavior scale have a correlation of -0.55. Corrected for attenuation, the true score correlation is -0.65. For parent ratings collected in Spring 2005, 2006, and 2007, the Social Competence and Problem Behavior scales have true-score correlations of -0.63, -0.65, and -0.65, respectively.

Table 3 displays the alpha reliabilities, concurrent correlations, and six-month longitudinal correlations for teacher and parent ratings of social competency and problem behavior. Longitudinal correlations show that ratings made by the same individual are fairly stable across time (i.e., from fall to spring). Also, teacher and parent ratings using the two factor solution show high internal consistency. Internal consistency coefficient alpha for teacher ratings using the Social Competency scale is 0.95 and 0.89 for the Problem Behavior scale. Internal consistency coefficient alpha for parent ratings using the Social Competency scale is 0.89 and 0.81 for the Problem Behavior scale.

However, mono-method correlations are higher than cross-informant correlations. For example, parent ratings of social competency correlate more strongly with parent ratings of problem behavior than they do with teacher ratings of social competency. While this indicates a lack of convergent validity, it is also not unexpected. De Los Reyes and Kazdin (2005) describe low cross-informant agreement (e.g., *r*s often in .20s) as “one of the most robust findings in clinical child research” (p. 483). These discrepancies have

Table 3

Concurrent and Six-Month Longitudinal Correlations for Teacher and Parent Reports of Social Competency and Problem Behavior

Time, rater, and scale			Fall 2004				Spring 2005			
			Teacher (*N* = 2320)		Parent (*N* = 2157)		Teacher (*N* = 2486)		Parent (*N* = 2096)	
			Social Competence	Problem Behavior	Social Competence	Problem Behavior	Social Competence	Problem Behavior	Social Competence	Problem Behavior
Fall 2004	Teacher	Social Competence	**.95**	-.64	.29	-.28	.71	-.52	.31	-.25
		Problem Behavior		**.89**	-.28	.34	-.53	.67	-.28	.29
	Parent	Social Competence			**.89**	-.55	.27	-.27	.68	-.40
		Problem Behavior				**.81**	-.28	.31	-.40	.60
Spring 2005	Teacher	Social Competence					**.95**	-.68	.32	-.28
		Problem Behavior						**.88**	-.31	.33
	Parent	Social Competence							**.90**	-.54
		Problem Behavior								**.82**

Note. Alpha reliabilities are shown in bold face on the diagonal.

been consistently found across methods for assessing behaviors in youths (e.g., rating scales, structured interviews, etc.) (e.g.,Gresham & Elliott, 1990; Renk & Phares, 2004).

While informant discrepancies have been well documented and replicated, research has generally not articulated and tested a theory to explain them. Explanations for this phenomenon have sometimes been attributed to differences in the contexts or situations in which different raters observe the child's behavior (Achenbach et al., 1987) and to different perceptions of what constitutes the behavior of interest. Currently, many researchers and clinicians regard reports from multiple informants as providing useful but different information about the child's behavior in different contexts, as viewed from different perspectives (De Los Reyes & Kazdin, 2005). Of course, to the extent that behavior ratings used in an efficacy study do not generalize across settings or contexts, outcomes of efficacy research using them should be interpreted with that limitation in mind. That is, an effect found for one type of rater should be interpreted within the bounds of the setting and not regarded as generalizable to other settings (or, generally speaking, traitlike).

Tables 4, 5, and 6 provide additional information about cross informant ratings using the SCRF. Across raters, ratings of social competency should correlate more highly with ratings of social competency than with ratings of problem behavior, and vice versa. These patterns hold true for both parent and teacher ratings of problem behavior, and for parent ratings of social competency. However, across raters and time, teacher ratings of social competency correlate just as highly with teacher ratings of problem behavior as they do with teacher ratings of social competency. In short, teacher ratings show essentially no evidence of discriminant validity, whereas parent ratings show modest

evidence of discriminant validity. Parents are generally, of course, the same rater at the different time points.

Table 4

Concurrent and Longitudinal Correlations for Teacher Reports of Social Competency and Problem Behavior at Three Points in Time

Rater and occasion		Teacher 1 (*N* = 2486)		Teacher 2 (*N* = 2577)		Teacher 3 (*N* = 2517)	
		SC[a]	PB[b]	SC	PB	SC	PB
Teacher 1	SC	1.00	-.68	.44	-.42	.45	-.44
	PB	(-.74)	1.00	-.41	.51	-.45	.54
Teacher 2	SC	(.46)	(-.45)	1.00	-.70	.48	-.45
	PB	(-.45)	(.57)	(-.76)	1.00	-.46	.52
Teacher 3	SC	(.47)	(-.49)	(.51)	(-.50)	1.00	-.69
	PB	(-.47)	(.60)	(-.48)	(.57)	(-.74)	1.00

Note. Teacher 1 is the student's classroom teacher in Spring 2005; Teacher 2 is the student's classroom teacher in Spring 2006; Teacher 3 is the student's classroom teacher in Spring 2007. Teachers 1, 2, and 3 are different individuals. Values included in parentheses below the diagonal are correlations corrected for attenuation.
[a] SC = Social Competency
[b] PB = Problem Behavior

Table 5

Concurrent and Longitudinal Correlations for Parent Reports of Social Competency and Problem Behavior at Three Points in Time

Occasion		Time 1 (*N* = 2096)		Time 2 (*N* = 1956)		Time 3 (*N* = 2110)	
		SC[a]	PB[b]	SC	PB	SC	PB
Time 1	SC	1.00	-.54	.64	-.36	.60	-.36
	PB	(-.63)	1.00	-.37	.57	-.34	.50
Time 2	SC	(.71)	(-.45)	1.00	-.56	.67	-.35
	PB	(-.42)	(.69)	(-.65)	1.00	-.38	.52
Time 3	SC	(.67)	(-.40)	(.74)	(-.44)	1.00	-.56
	PB	(-.42)	(.61)	(-.41)	(.63)	(-.65)	1.00

Note. Time 1 is the caregiver completing the rating scale in Spring 2005; Time 2 is the caregiver completing the rating scale in Spring 2006; Time 3 is the caregiver completing the rating scale in Spring 2007. Parent raters for Times 1, 2, and 3 are likely the same individual. Values included in parentheses below the diagonal are correlations corrected for attenuation.
[a] SC = Social Competency
[b] PB = Problem Behavior

Table 6

Cross Informant and True-Score Cross Informant Correlations

Rater and occasion		Parent, Time1		Parent, Time 2		Parent, Time 3	
		SC[a]	PB[b]	SC	PB	SC	PB
Teacher 1, Time 1	SC	.30	-.27	.27	-.24	.24	-.18
		(.32)	(-.31)	(.29)	(-.27)	(.26)	(-.20)
	PB	-.35	.35	-.31	.30	-.27	.24
		(-.39)	(.41)	(-.35)	(.35)	(-.30)	(.28)
Teacher 2, Time 2	SC	.28	-.26	.30	-.29	.32	-.23
		(.30)	(-.29)	(.32)	(-.33)	(.35)	(-.26)
	PB	-.25	.31	-.26	.31	-.25	.24
		(-.28)	(.36)	(-.29)	(.36)	(-.28)	(.28)
Teacher 3, Time 3	SC	.28	-.33	.28	-.29	.28	-.23
		(.30)	(-.37)	(.30)	(-.33)	(.30)	(-.26)
	PB	-.29	.36	-.27	.31	-.29	.27
		(-.32)	(.42)	(-.30)	(.36)	(-.32)	(.31)

Note. Rater 1 is the rater in Spring 2005; Rater 2 is rater in Spring 2006; Rater 3 is rater in Spring 2007. Values included in parentheses are correlations corrected for attenuation.
[a] SC = Social Competency
[b] PB = Problem Behavior

Covariates. Individuals' demographic characteristics and baseline scores were used as level-1 covariates. These include grade level, sex, ethnicity, FARM participation, and baseline parent or teacher ratings (Fall 2004). Due to the small percentage of ethnic minority students, ethnicity was recoded into a dichotomous variable. This grouping is not ideal, but there were too few students in most minority groups for meaningful subgroup analyses. White/Non-Hispanic and Asian/Pacific Islander students were grouped together as a historically academically advantaged group and African American, Hispanic, and Native American students were grouped together as a historically academically disadvantaged group. FARM participation was treated as a continuous variable, with 0 indicating students who did not qualify, 1 indicating students who qualified for reduced-cost meals, and 2 indicating students who qualified for free meals. Level-1 covariates were included to test for potential interaction effects of treatment with

individual characteristics. Table 7 displays the correlations between the level-1 covariates and the outcome variables.

Table 7

Correlations for Level-1 Covariates and Outcome Variables

Covariate	Teacher Outcome		Parent Outcome	
	SC [a]	PB [b]	SC	PB
Sex (1 = female)	.25	-.11	.20	-.06
Ethnicity (1 = academically advantaged)	.11	-.08	.06	-.02
FARM (1 = reduced meal, 2 = free meal)	-.14	.09	-.12	.09
Grade Level	.01	-.02	.06	.03
Teacher SC pretest	.41	(-.31)	(.22)	(-.10)
Teacher PB pretest	(-.40)	.35	(-.23)	(.17)
Parent SC pretest	(.27)	(-.19)	.59	(-.19)
Parent PB pretest	(-.35)	(.30)	(-.35)	.32

Note. Values in parentheses indicate correlations among variables not in the same model.
[a] SC = Social Competency
[b] PB = Problem Behavior

School characteristics were used as level-2 covariates to potentially increase power to detect a treatment effect. The covariates included were percentage of students mobile, which was available from the state's Department of Education website, and school aggregates of baseline parent or teacher ratings. Table 8 displays the correlations between the level-2 covariates and the aggregated outcome variables.

Table 8

School-Level Correlations for Level-2 Covariates and Aggregated Outcome Variables

Covariate	Aggregated Teacher Outcome		Aggregated Parent Outcome	
	SC [a]	PB [b]	SC	PB
Aggregate teacher pretest SC	.47	(-.31)	(.72)	(-.10)
Aggregate teacher pretest PB	(-.36)	.47	(-.59)	(.46)
Aggregate parent pretest SC	(.22)	(-.35)	.49	(-.13)
Aggregate parent pretest PB	(-.50)	(.31)	(-.56)	.71
Mobility	-.54	.75	-.57	.70

Note. Values in parentheses indicate correlations among variables not in the same model.
[a] SC = Social Competency; [b] PB = Problem Behavior

Analyses

Schools were assigned to treatment or control conditions. Students are nested within schools and cannot be treated as independent units of analysis. If students were to be treated as independent units, there is the risk of having unrealistically narrow confidence intervals and an increased likelihood of Type I errors. To avoid this pitfall, data were analyzed using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), with the individual student at level-1 and the school at level-2.

At baseline, the distributions of scores on parent and teacher ratings of problem behavior were highly skewed, with students being rated as displaying few problem behaviors. As a result, ratings of problem behavior were recoded into a dichotomous variable dividing the sample into a 'problem' group and a 'not problem' group.

Discussions of prevention and intervention estimate that 80 to 90 percent of students can be served by universal prevention strategies. The remaining 10 to 20 percent of students are considered to be at-risk or in need of intensive interventions (Sprague, Horner, & Walker, 2000). A recent sampling estimates the prevalence of Oppositional Defiant Disorder (ODD) to be at 10 percent (Nock, Kazdin, Hiripi, & Kessler, 2007). ODD is a diagnosis that represents an ongoing pattern of uncooperative, defiant, and hostile behavior and puts a child at risk for other mental disorders.

In light of these estimates, the distribution of parent and teacher reports of problem behavior were each recoded to indicate students who fell above and below the 90^{th} percentile, representing those students in the sample most at risk for significant behavior problems and in need of selected or intensive interventions. Students receiving scores in the 90^{th} percentile or above were coded 1 (i.e., the 'problem' group) and

students below the 90th percentile were coded as 0 (i.e., the 'not problem' group). Analyses were run to determine if treatment had an effect on a student's likelihood of belonging to the problem group. Hierarchical Generalized Linear Modeling (HGLM) using the Bernoulli sampling model using a logit link function was used to analyze the dichotomous outcome variable.

Parent and teacher ratings of social competency, although slightly skewed, fell along a more normal distribution. Ordinary HLM procedures were used to analyze this continuous outcome variable, which was standardized using a z-score to aid in interpretation of the coefficient.

Benchmark Analyses. The benchmark analyses are intended to estimate the effects of treatment for the sample of students who were present at the start of the study and at its completion three years later. These analyses include only those students who were originally assigned to the treatment condition in Fall 2004 and who should have received the total three years of treatment by Spring 2007 (i.e., the students in the first and second grades in Fall 2004 and in the third and fourth grades in Spring 2007). Students who moved out of the schools over the course of the three years are not included in these analyses, as they were not present for outcome data collection. There were a total of 836 students who had teacher ratings at both baseline and follow-up and a total of 683 students who had parent ratings at both baseline and follow-up. Focusing on the group of students who received the full dosage of treatment gives the program the best chance of demonstrating effects. The demographic characteristics of this sample are displayed in Table 9.

Table 9

Demographic Characteristics of Benchmark Sample

Sample	*N*	Percentage of Sample		
		Male	Academically Advantaged	Free or Reduced Meals
Cohort 1[a]				
Treatment	482	52	91	9
Control	412	52	87	9
Cohort 2[b]				
Treatment	503	47	92	12
Control	413	56	85	13

[a] Cohort 1 refers to those students who were in the first grade during the first year of implementation.
[b] Cohort 2 refers to those students who were in the second grade during the first year of implementation.

The following procedures were used for the analyses of ratings of social competency:

1) Fully Unconditional Model: A fully unconditional model is created to determine the proportion of the variance in the outcome that is between schools (i.e., intraclass correlation). No predictors are specified at any level in this model. In instances where the intraclass correlation is negligible, analysis of the data to assess possible treatment effects is not required unless there is an interest in cross level interactions.

$$Y_{ij} = \beta_{0j} + r_{ij} \quad (1)$$

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

Where

Y_{ij} = Behavior rating of student *i* in school *j*

β_{0j} = Average behavior rating in school *j*

2) Level-1 Model: The level-1 model reflects within-school regression. The level-1 model includes the following predictors: sex, ethnicity, FARM

participation, grade level, and baseline parent or teacher ratings (Fall 2004). Initially covariates were group-mean centered and the error terms at level-2 were freed in order to determine whether they had fixed or random effects on the outcome variable. Then all variables were entered as grand-mean centered unless slopes were random across schools. Slopes were fixed unless it was determined that the slope had statistically significant random variance. If the slope had significant random variance, the interaction of treatment with the demographic characteristic was tested with that covariate group-mean centered. However, there were no significant treatment interactions with level-1 covariates. Because potential treatment interactions were a subsidiary area of interest and the subgroup sample sizes were relatively small for examining this issue, all level-1 slopes were ultimately treated as fixed in order to conserve degrees of freedom. All covariates were retained in the level-1 model irrespective of whether or not they were statistically significant in explaining variance in the outcome.

Level One (Student)

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \overline{X}_{1\cdot\cdot}) + \beta_{2j}(X_{2ij} - \overline{X}_{2\cdot\cdot}) + \beta_{3j}(X_{3ij} - \overline{X}_{3\cdot\cdot}) \qquad (1)$$

$$+ \beta_{4j}(X_{4ij} - \overline{X}_{4\cdot\cdot}) + \beta_{5j}(X_{5ij} - \overline{X}_{5\cdot\cdot}) + r_{ij}$$

Where

Y_{ij} = Behavior rating of student *i* in school *j*

β_{0j} = Average behavior rating in school *j*

β_{1j} = Regression slope for sex covariate

$(X_{1ij} - \overline{X}_{1\cdot\cdot})$ = Sex covariate grand-mean centered

β_{2j} = Regression slope for ethnicity covariate

$(X_{2ij} - \overline{X}_{2\cdot\cdot})$ = Ethnicity covariate grand-mean centered

β_{3j} = Regression slope for the FARM covariate

$(X_{3ij} - \overline{X}_{3\cdot\cdot})$ = FARM covariate grand-mean centered

β_{4j} = Regression slope for grade level covariate

$(X_{4ij} - \overline{X}_{4\cdot\cdot})$ = Grade level covariate grand-mean centered

β_{5j} = Regression slope for baseline behavior rating covariate

$(X_{5ij} - \overline{X}_{5\cdot\cdot})$ = Baseline behavior rating covariate grand-mean centered

r_{ij} = Error term associated with individual *i* in school *j*

3) Level-2 Model: The level-2 model reflects between-school variance. Treatment (i.e., whether or not the school implemented Second Step), percentage of students mobile, and school aggregates of baseline parent or teacher ratings were included at level-2. Percentage of students mobile and baseline aggregates were entered as grand-mean centered. The treatment variable was entered as uncentered, with 0 indicating the control group schools. The coefficient for treatment is the estimate of the treatment effect on the outcome variable. Mobility and aggregate baseline covariates were included to potentially increase power to detect a treatment effect. As displayed in Table 8, student mobility was highly correlated with the outcome variables (r = -0.5 for social competency; r = 0.7 for problem behavior), and school-level pretest scores, when added to individual-level pretest scores, have the potential to reduce the minimum detectable effect size (Bloom, Richburg-Hayes, & Black, 2007). Due to the lack of significant treatment

interactions at level-1 and the few degrees of freedom at level-2, treatment interactions were not explored at the school-level in order to conserve degrees of freedom.

Level Two (School)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}(W_{2j} - \overline{W}_{2\cdot}) + \gamma_{03}(W_{3j} - \overline{W}_{3\cdot}) + u_{0j} \quad (1)$$

$$\beta_1 = \gamma_{10} \quad (2)$$

$$\beta_2 = \gamma_{20} \quad (3)$$

$$\beta_3 = \gamma_{30} \quad (4)$$

$$\beta_4 = \gamma_{40} \quad (5)$$

$$\beta_5 = \gamma_{50} \quad (6)$$

Where

γ_{00} = Average behavior rating across schools

γ_{01} = Effect parameter for treatment

W_{1j} = Treatment indicator (0 = control, 1 = treatment)

γ_{02} = Effect parameter for aggregate baseline according to my model

$(W_{2j} - \overline{W}_{2\cdot})$ = Aggregate baseline (z-score), grand-mean centered

γ_{03} = Effect parameter for mobility according to my model

$(W_{3j} - \overline{W}_{3\cdot})$ = School mobility (percentage), grand-mean centered

u_{0j} = Error term associated with school *j*

β_1 = Average within-school slope in the regression of outcome behavior rating on student sex

β_2 = Average within-school slope in the regression of outcome behavior rating on student ethnicity

β_3 = Average within-school slope in the regression of outcome behavior rating on student FARM status

β_4 = Average within-school slope in the regression of outcome behavior rating on grade level

β_5 = Average within-school slope in the regression of outcome behavior rating on baseline behavior rating

The following models using a dichotomous outcome variable were created for the analyses of ratings of problem behavior:

1) Level-1 Model: The level-1 model estimates the difference in the log-odds of being in the problem group between two students who attend the same school but differ by one unit on the level-1 predictor.

Level One (Student)

$$\eta_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \overline{X}_{1\cdot\cdot}) + \beta_{2j}(X_{2ij} - \overline{X}_{2\cdot\cdot}) + \beta_{3j}(X_{3ij} - \overline{X}_{3\cdot\cdot}) \quad (1)$$

$$+ \beta_{4j}(X_{4ij} - \overline{X}_{4\cdot\cdot}) + \beta_{5j}(X_{5ij} - \overline{X}_{5\cdot\cdot})$$

Where

η_{ij} = Log of the odds of being in the problem group

β_{1j} = Regression slope of student sex on the log odds of being in the problem group

$(X_{1ij} - \overline{X}_{1\cdot\cdot})$ = Sex covariate grand-mean centered

β_{2j} = Regression slope of student ethnicity on the log odds of being in the problem group

$(X_{2ij} - \overline{X}_{2\cdot\cdot})$ = Ethnicity covariate grand-mean centered

β_{3j} = Regression slope of student FARMS status on the log odds of being in the problem group

$(X_{3ij} - \overline{X}_{3\cdot\cdot})$ = FARM covariate grand-mean centered

β_{4j} = Regression slope of student grade level on the log odds of being in the problem group

$(X_{4ij} - \overline{X}_{4\cdot\cdot})$ = Grade level covariate grand-mean centered

β_{5j} = Regression slope of student baseline behavior ratings on the log odds of being in the problem group

$(X_{5ij} - \overline{X}_{5\cdot\cdot})$ = Baseline behavior rating covariate grand-mean centered

2) Level-2 Model:

Level Two (School)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}(W_{2j} - \overline{W}_{2\cdot}) + \gamma_{03}(W_{3j} - \overline{W}_{3\cdot}) + u_{0j} \quad (1)$$

$$\beta_1 = \gamma_{10} \quad (2)$$

$$\beta_2 = \gamma_{20} \quad (3)$$

$$\beta_3 = \gamma_{30} \quad (4)$$

$$\beta_4 = \gamma_{40} \quad (5)$$

$$\beta_5 = \gamma_{50} \quad (6)$$

Where

β_{0j} = Covariate adjusted odds of being in the problem group for school *j*

γ_{01} = Effect parameter for treatment

W_{1j} = Treatment indicator (0 = control, 1 = treatment)

γ_{02} = Effect parameter for aggregate baseline according to my model

$(W_{2j} - \overline{W}_{2\cdot})$ = Aggregate baseline (z-score), grand-mean centered

γ_{03} = Effect parameter for mobility according to my model

$(W_{3j} - \overline{W}_{3\cdot})$ = School mobility (percentage), grand-mean centered

u_{0j} = Error term associated with school *j*

Sensitivity Analyses. Weighted and imputed analyses were conducted to cope with issues of attrition.

Weighted Analyses. Students in the sample from academically disadvantaged ethnic groups and students eligible to participate in the FARM program were less likely to participate in the study. In order to better represent these groups, analyses using weights were conducted. Cells were created by stratifying the sample by school, ethnicity, and FARM status. Weights are based on the within group response rate for each subgroup, and were created by calculating the inverse of the probability that cases in each cell had a completed parent or teacher rating. Higher weight values were given to minority respondents. This procedure will reduce bias if complete-case individuals differ from missing data individuals from the same subgroup completely at random.

Imputed Analyses. Students were lost from the sample due to either non-consent or withdrawal from the school prior to the end of the study. The imputed analyses used an intent-to-treat sample, which included all students present at the start of the study regardless if they had outcome data in Spring 2007 (N = 1810). An Expectancy Maximization (EM) algorithm was used to impute data that were missing due to either non-consent or withdrawal from the study prior to the final Spring 2007 data collection. The maximum-likelihood method uses available information (i.e., demographic information, school membership, etc.) to replace missing data with predicted values. Participants are treated as if they received a full dose of the treatment, regardless of

whether or not they actually did. This analysis provides an unbiased estimate of the effects of being assigned to the treatment condition if data are missing at random (Rubin, 1996).

Results

Social Competency

Tables 10 and 11 display the effects for the level-1 and combined models of teacher and parent ratings of student social competency. The ICC values indicate the amount of between school variance in the dependent variable. For both teacher and parent ratings, less than five percent of the variance was between schools. Treatment did not have a statistically significant effect on ratings of social competency. Treatment coefficients (γ_{01}) were near zero for teacher ratings of social competency (coefficient = 0.02, *SE* = 0.15) and parent ratings of social competency (coefficient = -0.04, *SE* = 0.08). Students in treatment schools, net of individual characteristics, were not rated as more or less socially competent by teachers or parents after three years of program implementation. For both models, sex (coefficient = 0.31, *SE* = 0.06), FARM status (coefficient = -0.17, *SE* = 0.07), and baseline ratings (coefficient = 0.37, *SE* = 0.03) were significant level-1 predictors. Male students, students participating in the FARM program, and students with lower social competency ratings at baseline were rated as less socially competent three years later. No level-2 covariates reached statistical significance; aggregated baseline ratings and percentage of students mobile were not predictive of individual student social competency. In other words, school contextual factors examined did not explain variance in the outcome above and beyond individual student variables. For teacher ratings of social competency, treatment interactions were tested for sex (treatment coefficient = 0.11, *SE* = 0.24), pretest (treatment coefficient = -0.07, *SE* = 0.11), and grade-level (treatment coefficient = -.18, *SE* = 0.24), but none were found to

be significant. No treatment interactions were tested for parent ratings of social competency, as no level-1 covariates had randomly varying slopes.

Table 10

Level-One Model and Combined Model of Teacher Ratings of Student Social Competency, ICC = 0.04

Fixed effects	Level-1 Model			Combined Model		
Variable	Coefficient (*SE*)		*df*	Coefficient (*SE*)		*df*
Intercept, γ_{00}	-0.01	(0.06)	11	-0.03	(0.10)	8
Individual sex (0 = male, 1 = female), γ_{10}	0.31*	(0.06)	830	0.31*	(0.06)	827
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian), γ_{20}	0.17	(0.14)	830	0.16	(0.14)	827
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch), γ_{30}	-0.17*	(0.07)	830	-0.15*	(0.08)	827
Grade level, γ_{40}	-0.02	(0.06)	830	-0.02	(0.06)	827
Fall 2004 baseline teacher rating, γ_{50}	0.37*	(0.03)	830	0.37*	(0.03)	827
Treatment status (0 = control, 1 = treatment), γ_{01}	---	---	---	0.02	(0.15)	8
Aggregated baseline teacher rating, γ_{02}	---	---	---	-0.06	(0.08)	8
Mobility, γ_{03}	---	---	---	-0.02	(0.02)	8

Note. SE = standard error
*$p < .05$

Table 11

Level-One Model and Combined Model of Parent Ratings of Student Social Competency, ICC = 0.01

Fixed effects	Level-1 Model			Combined Model		
Variable	Coefficient (*SE*)		*df*	Coefficient (*SE*)		*df*
Intercept	-0.01	(0.04)	11	0.00	(0.06)	8
Individual sex (0 = male, 1 = female), γ_{10}	0.18*	(0.06)	677	0.18*	(0.06)	674
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian), γ_{20}	0.04	(0.16)	677	0.02	(0.17)	674
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch), γ_{30}	-0.18*	(0.09)	677	-0.18*	(0.06)	674
Grade level, γ_{40}	0.01	(0.06)	677	0.01	(0.06)	674
Fall 2004 baseline parent rating, γ_{50}	0.59*	(0.03)	677	0.59*	(0.03)	674
Treatment status (0 = control, 1 = treatment), γ_{01}	---	---	---	-0.04	(0.08)	8
Aggregated baseline parent rating, γ_{02}	---	---	---	0.06	(0.04)	8
Mobility, γ_{03}	---	---	---	-0.00	(0.01)	8

Note. SE = standard error.
*$p < .05$

To cope with missing data, sensitivity analyses were run using weights based on school, ethnicity, and FARM status and imputation using an EM algorithm. Weighting

reconstitutes the sample if data for people within weighting subgroups are missing completely at random, and imputation re-constitutes the sample if data for people with missing data are missing at random given other available data about them. Results from the sensitivity analyses support those obtained in the primary analyses. Table 12 displays the treatment coefficients for primary and sensitivity analyses. For both parent and teacher ratings of social competency, coefficients consistently hover near zero and are not statistically significant.

Table 12

Treatment Effect on Ratings of Social Competency Ratings, Primary and Sensitivity Analyses

Rater	Benchmark			Weighted			Imputed		
	Coefficient	*SE*	*p*	Coefficient	*SE*	*p*	Coefficient	*SE*	*p*
Teacher	0.02	0.15	0.88	0.02	0.02	0.90	-0.04	0.09	0.63
Parent	-0.04	0.08	0.57	-0.06	0.06	0.32	0.01	0.07	0.92

Problem Behavior

Tables 13 and 14 display the effects for the level-1 and combined models of teacher and parent ratings of student problem behavior. The distribution of parent and teacher reports of problem behavior were each recoded to indicate students who fell above and below the 90th percentile, representing those students in the sample most at risk for significant behavior problems and in need of selected or intensive interventions. Analyses were run to determine if treatment had an effect on a student's likelihood of belonging to the problem group. The results were inconsistent for teacher and parent reports of problem behavior. For teacher ratings, students in the treatment schools were less likely to be rated in the problem range than were students in the control schools (odds ratio = 0.67, *CI* = 0.34 – 1.33). This is a statistically nonsignificant result, and

plausible effect estimates include both increases and decreases in the odds of being classified a “problem” child. The odds that parents of students in treatment schools would rate students in the “problem” child range were almost twice those of parents of students in the control schools (odds ratio = 1.93, *CI* = 0.98- 3.80). This result was also statistically nonsignificant, although only by a small margin; the confidence interval included a change in odds of 1.0. For teacher ratings, FARM status (odds ratio = 1.66, *CI* = 1.07 – 2.57) and baseline ratings (odds ratio = 6.71, *CI* = 4.13 – 10.90) were significant level-1 predictors; for parent ratings, baseline problem behavior (odds ratio = 15.32, *CI* = 7.16 – 32.81) was the only significant level-1 predictor. Students participating in the FARM program and students with higher problem behavior ratings at baseline were more likely to be rated in the problem group by teachers, and students with higher problem behavior ratings at baseline were more likely to be rated in the problem group by parents. No level-2 predictors were statistically significant in either model; aggregated baseline ratings and percentage of students mobile were not predictive of whether or not an individual was rated in the problem group. Again, school contextual factors did not explain variance in the outcome above and beyond individual student variables. For parent ratings of problem behavior, treatment interactions were tested for pretest (odds ratio = 0.47, *CI* = 0.03 – 7.02) and grade-level (odds ratio = 0.22, *CI* = 0.03 – 1.95), but neither was found to be significant. No treatment interactions were tested for teacher ratings of problem behavior, as no level-1 covariates had randomly varying slopes.

Sensitivity analyses using weights and imputation produced results consistent with those found in the primary analyses. Teachers in treatment schools were less likely to rate students as displaying behavior problems in the problem range, but to a

Table 13

Level-One Model and Combined Model of Teacher Ratings of Problem Behavior

Fixed effects	Level-1 Model					Combined Model				
Variable	Coefficient (*SE*)		*df*	*OR*	(*CI*)	Coefficient (*SE*)		*df*	*OR*	(*CI*)
Intercept, γ_{00}	-2.21	(0.16)	11	0.11	(0.08 – 0.16)	-2.01	(0.22)	8	0.13	(0.08 - 0.22)
Individual sex (0 = male, 1 = female), γ_{10}	-0.35	(0.24)	830	0.70	(0.44 – 1.13)	-0.35	(0.24)	827	0.70	(0.44 - 1.13)
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian), γ_{20}	-0.26	(0.42)	830	0.77	(0.34 – 1.76)	-0.25	(0.42)	827	0.78	(0.34- 1.78)
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch), γ_{30}	0.52*	(0.21)	830	1.69*	(1.12 – 2.55)	0.50*	(0.22)	827	1.66*	(1.07 - 2.57)
Grade level, γ_{40}	-0.05	(0.23)	830	0.95	(0.61 – 1.50)	-0.05	(0.23)	827	0.95	(0.60 - 1.50)
Fall 2004 baseline teacher rating, γ_{50}	1.87*	(0.24)	830	6.48*	(4.04 – 10.40)	1.90*	(0.25)	827	6.71*	(4.13 - 10.90)
Treatment status (0 = control, 1 = treatment), γ_{01}	---	---	---	---	---	-0.40	(0.31)	8	0.67	(0.34 - 1.33)
Aggregated baseline teacher rating, γ_{02}	---	---	---	---	---	-0.20	(0.19)	8	0.82	(0.54 - 1.24)
Mobility, γ_{03}	---	---	---	---	---	0.04	(0.04)	8	1.04	(0.95 - 1.14)

Note. SE = Standard error; *OR* = Odds ratio.
$^{*}p < .05$

Table 14

Level-One Model and Combined Model of Parent Ratings of Student Problem Behavior

Fixed effects	Level-1 Model					Combined Model				
Variable	Coefficient (*SE*)		*df*	*OR*	(*CI*)	Coefficient (*SE*)		*df*	*OR*	*(C1)*
Intercept	-2.48	(0.17)	11	0.08	(0.06 – 0.12)	-2.89	(0.28)	8	0.06	(0.03 - 0.10)
Individual sex (0 = male, 1 = female), γ_{10}	-0.30	(0.28)	677	0.74	(0.42 – 1.28)	-0.37	(0.29)	674	0.69	(0.40 - 1.21)
Ethnicity (0 = African American, Hispanic, Native American; 1 = White, Asian), γ_{20}	0.35	(0.81)	677	1.42	(0.29 – 6.89)	0.46	(0.83)	674	1.59	(0.31 - 8.08)
FARM status (0 = no FARMS, 1 = reduced lunch, 2 = free lunch), γ_{30}	0.08	(0.33)	677	1.08	(0.56 – 2.07)	0.06	(0.33)	674	1.06	(0.54 - 2.08)
Grade level, γ_{40}	0.38	(0.28)	677	1.46	(0.85 – 2.53)	0.39	(0.28)	674	1.47	(0.85 - 2.56)
Fall 2004 baseline parent rating, γ_{50}	2.62*	(0.38)	677	13.80*	(6.61 – 28.79)	2.73*	(0.39)	674	15.32*	(7.16 - 32.81)
Treatment status (0 = control, 1 = treatment), γ_{01}	---	---	---	---	---	0.66	(0.31)	8	1.93	(0.98 - 3.80)
Aggregated baseline parent rating, γ_{02}	---	---	---	---	---	0.06	(0.16)	8	1.06	(0.75 - 1.51)
Mobility, γ_{03}	---	---	---	---	---	0.02	(0.05)	8	1.02	(0.92 - 1.13)

Note. SE = Standard error; *OR* = Odds ratio; $^{*}p < .05$

Table 15

Treatment Effect on Ratings of Problem Behavior, Primary and Sensitivity Analyses

Rater	Benchmark			Weighted			Imputed			Simplified		
	OR	*CI*	*p*	*OR*	*CI*	*p*	*OR*	*CI*	*p*	*OR*	*CI*	*p*
Teacher	0.67	0.34 - 1.33	0.23	0.73	0.40 - 1.32	0.28	0.82	0.48 - 1.41	0.45	0.63	0.32 - 1.25	0.17
Parent	1.93	0.98 - 3.80	0.07	2.01*	1.19 - 3.41	0.02	1.78*	1.09 - 2.90	0.03	2.05*	1.06 - 3.99	0.04

Note. OR = Odds ratio; *CI* = Confidence interval

$^{*}p < .05$

statistically nonsignificant extent across all three analyses. Parents of students in treatment schools were approximately two times more likely to rate their child in the problem range. These coefficients reached statistical significance in the weighted and imputed sensitivity analyses. Table 15 displays the treatment coefficients for primary and sensitivity analyses.

When comparing the level-1 and combined models for problem behavior, it was observed that some of the coefficients and standard errors slightly increased, suggesting minor multicollinearity in the model. In order to determine if multicollinearity was affecting the size and significance of coefficients, a very simplified model was tested as an additional sensitivity analysis. This model included baseline ratings as the only covariate at level-1 and treatment as the only covariate at level-2. Results obtained in the simplified model for teacher ratings of problem behavior (odds ratio = 0.63, *CI* = 0.32 – 1.25) and for parents ratings of problem behavior (odds ratio = 2.05, *CI* = 1.06 – 3.99) are approximately equal to those found in the more complex model, indicating that the coefficients are not inflated as a result of multicollinearity.

The coefficients for parent ratings of problem behavior were the only ones to reach statistical significance, and the results indicate a negative treatment effect. Only one of the prior studies reviewed (Grossman et al., 1997) measured parent perceptions of student behavior, so little is known about how the Second Step curriculum may influence parent ratings. The data were further explored to better understand the result. Frequency counts of the number of problem-range students were calculated by school (see Appendix D). Treatment and control schools look similar in the in the number of students in the problem range at pretest. At posttest, the number of students in the problem group in the

control schools decreased, whereas the number of students in the problem group in the treatment schools stayed about the same as at pretest. An examination of the frequency counts also showed that there was one outlier school. Thirty percent of the students in this school were rated in the problem range at posttest. The analyses were re-run without that outlier school and its matched pair to determine if the outlier was driving the results. When the outlier school and its matched pair are omitted, the negative treatment effect is almost significant (odds ratio = 2.11, *CI* = 1.01 - 4.38, $p = 0.06$).

Discussion

Second Step is a widely used violence prevention curriculum that is marketed as a research-supported program based on weak evidence derived from the studies reviewed earlier in this paper. The studies reviewed generally demonstrated weak non-experimental designs, single year trials, and a failure to measure implementation fidelity. This study sought to improve upon the existing evaluations of Second Step by conducting a randomized control group trial using maximally sensitive measures collected following three years of full, measured program implementation.

Results indicated no positive effects of the Second Step curriculum on parent and teacher perceptions of student social competency or problem behavior. Specifically, no evidence for the efficacy of the Second Step program in promoting social competencies was found when the analysis focused on the subset of competency items that were selected because these items may be particularly sensitive to the effects of the program. Moreover, there were no significant effects for parent and teacher reports of social competency in the sensitivity analyses. For problem behavior, teachers were less likely to report students in the treatment schools in the problem range; however, this effect did not approach significance. Parent reports of problem behavior showed a negative effect for the treatment condition in some, but not all, analyses. Participation in the Second Step program for three years almost doubled the odds of being rated in the problem range by parents, an effect that was marginally statistically significant in some analyses and nonsignificant in others.

The prior analyses of SCRF ratings as a single scale concluded that Second Step had no significant effects on parent or teacher behavior ratings (Gottfredson et al., 2008).

These two studies and the Grossman et al. (1997) study, the only true experimental evaluations of Second Step to date, provide convergent evidence that Second Step does not have important positive effects on parent and teacher perceptions of student behavior.

Some positive effects of Second Step have been reported in these two studies for direct observational measures of student behavior. Grossman et al. (1997) found significant effects for reductions in physical aggression and increases in neutral/prosocial behavior, and no significant effects for verbal aggression. These effects were found in the cafeteria and on the playground, but not in the classroom. Cooke et al. (2007), however, using the same observational procedures, was unable to replicate these findings. Gottfredson et al. (2008) found significant increases in observed prosocial behaviors during a contrived team-work activity in the classroom, but also found statistically nonsignificant increases in disengagement, verbal aggression, and physical aggression.

Subtle behavior changes may be more difficult for parents or teachers to detect than for trained observers who are charged specifically with recording student behavior. However, it is also important to consider the meaningfulness of behavior changes that are not detectable to parents, teachers, or even the children themselves. Also, reported interrater reliability for the behaviors observed was not at an acceptable level. Grossman et al. (1997) had fair interrater reliability for physical ($\kappa = 0.50$) and verbal aggression ($\kappa = 0.45$), and high reliability for neutral/prosocial behavior ($\kappa = 0.92$) but only after collapsing the two categories into one because raters were not able to distinguish between neutral and prosocial behaviors. Using the same observational procedures as Grossman, Cooke et al. (2007) did not establish acceptable reliability for neutral/positive verbal behaviors, but had acceptable reliability for the other behaviors recorded. Gottfredson et

al. (2008) reported interrater correlation coefficients ranging from .51 to 1.0. If potential social skills program effects are best detected by behavioral observations, future research is needed to develop behavior observation protocols that are able to reliably capture the behaviors of interest and to richly reflect student behavior in populations that may display a low base rate of problem behaviors.

This study's results diverge from meta-analytic studies that show positive effects of prevention programs on increasing student social competency and decreasing problem behavior. It is possible that the general results may not directly apply to the low-risk population included in this study. Wilson and Lipsey (2007) report that, for the studies included in their meta-analysis, the participants receiving universal interventions could actually be considered at-risk populations due to the prevalence of low socioeconomic or high crime neighborhoods. Durlak and Wells (1997) combined universal and high-risk selected populations when running analyses. In the analysis conducted by Hahn et al. (2007), 35% of the studies conducted at the elementary school level were in low SES or high crime neighborhoods, with results suggesting larger reductions in aggression in low SES/high crime populations. The schools participating in this study were generally well-functioning and located in middle-class, low-crime neighborhoods. There were relatively few students from low socioeconomic backgrounds, and while no treatment interactions were found with this subset of students, the sample was too small for fully exploring interaction effects. The possibility that primary prevention programs are more effective for high-risk versus low-risk populations should continue to be investigated.

While ecological variables such as neighborhood context and family environment have been found to add significantly to the prediction of parent- and teacher-rated

behavioral outcomes (Greenberg, Coie, Lengua, & Pinderhughes, 1999), school context covariates were not significant predictors of student behavior in this study. Highly predictive level-2 covariates (i.e., aggregated baseline ratings and percentage of students mobile) were no longer predictive of individual student social competency or problem behavior once the student characteristics were controlled for at level-1. Once individual-level variables were controlled for, school-level variables were no longer associated with the outcome. This suggests that individual student characteristics, not the school ecology, are explaining individual student behavior, as rated by teachers and parents. Also, boys, students from low socio-economic backgrounds, and students rated at baseline as displaying fewer social skills and more problem behaviors were rated three years later as displaying fewer social skills and more problem behaviors than their peers. While prevention programs have been found to be more effective for younger students, students with low socioeconomic status (Wilson & Lipsey, 2007), and students with higher baseline aggression (Kellam, Rebok, Ialongo, & Mayer, 1994), no significant treatment interactions with student demographic variables were found. It is possible that some packaged program curricula are not well-matched to the school and home needs of those subgroups of children who consistently experience more discipline problems in schools and the community. However, Second Step was not more or less effective for specific subgroups of students.

The negative effect on parent ratings of problem behavior, the only effect to reach statistical significance in some analyses, may be an example of fishing and the error rate problem. The probability of a Type 1 error (i.e., finding statistically significant differences in sample data when these differences do not exist in the population)

increases when multiple statistical tests are computed on the same data set (Shadish, Cook, and Campbell, 2002). Between the original set of analyses (Gottfredson et al., 2008) and this study's subsequent analyses, these data have now been used in 12 primary and 24 sensitivity analyses, thereby increasing the probability of obtaining a statistically "significant" result simply by chance. When considered in context with the three other nonsignificant effects found in this study, it would be premature to conclude that the Second Step program has a harmful effect on students. This is particularly so when the hypothesis for the present analyses was that the social competency items and not the behavior problem items would be more sensitive to treatment effects for this program.

Implications for Current Practice and Future Research

These results are an important contribution to the research on school-based prevention programs, particularly in today's educational climate, which enacts policy and directs monies toward evidence-based intervention efforts in the schools. While meta-analytic studies indicate that prevention efforts are generally effective, strong evaluations of specific programs are rare. Furthermore, of the large-scale evaluations conducted, most are conducted by the program developers. In order for scientific research to inform policy decisions, more independent, methodologically sound evaluations of school programs should be conducted. This study responded to that need by conducting a randomized, multi-year evaluation of a widely used school-based prevention program. Even when evaluated under favorable conditions (i.e., well-implemented, multi-year, maximally sensitive measures), evidence is accumulating that this widely used intervention program is not effective in improving student outcomes. Such knowledge should be used when school administrators are considering where to invest their time and resources.

There are implications of this study on the measurement of student behavior in the context of prevention research. Parent and teacher judgments are important socially valid indicators of the appropriateness of student behavior (Kellam, Branch, Agrawal, & Ensminger, 1975), and are therefore meaningful outcomes to consider when studying the effectiveness of prevention programs. Also, gathering data from multiple raters in different settings is important in painting a more comprehensive picture of a child, as well as for assessing the effectiveness of interventions across settings (Gresham & Elliott, 1990). Many studies collect teacher ratings of school-based prevention programs, but far fewer collect ratings from parents. Second Step lessons are delivered in the classroom, but the curriculum also includes homework and strategies aimed at generalization of skills across settings. As such, it is important to measure potential effects both at school and at home. This study's finding of a possible null effect or negative effect for parent ratings further reinforces the importance of measuring the influence of school-based programs on behavior at home.

Low cross-informant agreement has been characterized as "one of the most robust findings in clinical child research" (De Los Reyes & Kazdin, 2005, p. 483). The repeated demonstration of cross-informant discrepancies raises concerns about the reliability of any single source of information and about what is actually being measured. It is possible that social competency and problem behavior are so situational that the behaviors do not transcend settings. Or, more likely, that variance due to rater is so great that the measure narrowly reflects teacher and parent perceptions of behavior rather than an overarching trait. Furthermore, raters tend to make global impressions of related traits, and ratings scales therefore do not finely discriminate between traits. Across raters and time, teacher

ratings using the SCRF failed to show discriminant validity between ratings of social competency and problem behavior. Parent ratings showed modest evidence of discriminant validity, but these are typically the same rater for each occasion. The failure to demonstrate convergent and discriminant validity using the SCRF raises doubt that this tool is able to adequately distinguish between social competency and problem behavior.

A review of the literature on prevention shows that risk and protective factors are pervasively listed as distinct, and that evaluation studies routinely measure these outcomes separately. However, there is little empirical exploration of whether these are two distinct constructs. The measure used in this study failed to demonstrate adequate convergent and discriminant validity across raters and scales. However, the SCRF is a measure adapted and modified for the purpose of evaluating the Second Step program, and was not designed for the purpose of assessing the matter of risk versus protection. Future psychometric papers should target this topic using measures and analysis procedures specifically designed for that purpose. The prevention field would benefit from an empirical evaluation of this theory before it continues to perpetuate the two-pronged conceptualization of risk.

Limitations

As with any research project, there are areas of limitation. First, although the total number of participants is high, schools, not students, were assigned to the treatment condition. Having only 12 schools restricted the possibility of testing for treatment interactions at level-2. It remains unexplored how school environmental factors may interact with treatment in affecting individual students. Also, with only 12 schools, the power to detect effects is relatively low. However, while the power to detect effects is

low, the nonsignificant coefficients were near zero, so even with improved power, any newly statistically significant effects would not be practically meaningful.

This study may also have been limited by the homogeneity of the population. The small number of students within subgroups restricted the opportunity to examine differential effectiveness for subgroups. Additionally, implementation data reflect the percentage of Second Step lessons delivered by the classroom teacher, and not necessarily the percentage of lessons received by each student. Lesson delivery may have overlapped with pull-out services such as ESOL, special education, or Title 1 services. If this is true, the students most at risk may not have been fully exposed to the intervention, further restricting the possibility of detecting treatment interactions with specific student subgroups. Also, attrition analyses revealed that students from academically disadvantaged ethnic groups and students participating in the FARM program were less likely to have completed data in the third year of the study, which restricts the generalizability of the results. While certain subgroups were less likely to participate in the study in general, there were no interactions with treatment, indicating that these students were less likely to participate in the study to an equivalent degree in treatment and control schools. Attrition, therefore, does not restrict internal validity. Sensitivity analyses using weighting and imputation were conducted to reconstruct the sample if students were missing at random within schools, ethnic groups, and FARM status. The results from these sensitivity analyses replicate the findings from the primary analyses.

Implementation data indicate that Second Step required a higher frequency and formality of instruction and more formal materials than the activities being implemented in the control schools. The schools involved in this study reported a high level of

implementation of the full lessons and their critical components. D. Gottfredson, G. Gottfredson, and Skroban (1998) stated that schools in the most need of prevention efforts are also those least equipped to meet the high-quality implementation demands associated with positive program effects. Considering the high amount of organization required to fully implement the Second Step program, it is possible that these results do not generalize to schools that do not have the resources and organization to implement such a structured program. Also, implementation was measured using a teacher self-report form on which teachers indicated if they implemented the lesson and each of its critical components. These ratings are subject to the teachers' perceptions of their own performance and were not corroborated by an independent observer. Furthermore, the ratings indicate whether the lesson and components were delivered, not whether they were delivered well. Considering the enthusiasm and technical support provided by the school administrators, counselors, and research staff in promoting program implementation, it is expected that the quality of implementation was higher than would be met at a typical school. Still, implementation information would be improved had individual variation in the quality of implementation been directly assessed.

Implementation data show that while control schools were implementing various character education initiatives, treatment schools demonstrated distinct differences in the time spent in formal character education instruction and in the delivery of specific program features intended by the Second Step curriculum (e.g., modeling, role play, etc.). When compared to this control group, despite the noted limitations, this study offers compelling evidence that Second Step is not effective in improving parent and teacher perceptions of student behavior.

Appendix A

Table A1

Completion Rates for Second Step Curriculum by Year

Second Step curriculum	Year 1 (*N* = 111)	Year 2 (*N* = 109)		Year 3 (*N* = 113)	
	Percent	Percent	No Modification	Percent	No Modification
Lessons Completed	91	96	—	91	—
Activities					
Modeling	86	91	81	85	83
Role-play	85	89	80	83	83
Video	87	94	97	92	95
Distribution	62	69	82	65	84
List-making	79	91	72	87	75
Evaluation	79	84	71	78	77

Note. "Percent" indicates the percentage of lessons completed. "No Modification" indicates the percentage of activities performed that were implemented as prescribed by the Second Step curriculum (i.e., without modification). The modification measure was not added to the implementation logs until the second year of program implementation.

Appendix A

Table A2

Teacher Report of Social and Character Education Materials Used and Activities Delivered in the Classroom

	Treatment			Control		
Variables	Mean	*SD*	*N*[a]	Mean	*SD*	*N*
Materials						
Teacher or instructor guides 0 = no manual, 2 = detailed manual	1.89	.32	45	.68	.75	31
Materials for implementation 0 = no ready-made materials, 3 = all materials provided	2.49	.70	45	.68	.83	31
Materials distributed to students 0 = no ready-made materials, 3 = all materials provided	1.26	.66	43	.33	.66	30
Activities						
Formality of instruction 0 = very informal, 3 = very formal	1.11	.76	104	.44	.69	72
Frequency of formal instruction 0 = < once a month, 2 = once a week, 5 = once a day or more	3.17	.69	108	1.77	1.13	81
Time specifically allotted 0 = not true, 2 = very true[b]	1.48	.69	108	1.24	.78	80
Instruction is required	1.93	.30	108	1.04	.76	79
Principal support	1.87	.34	108	1.56	.57	78
Learning points shared with students early in the lesson	1.61	.60	106	1.05	.69	76
Students summarize and record learning points	1.12	.71	106	.71	.74	78
Student discussion of realistic social situations	1.89	.34	108	1.46	.62	78
Instructor names and models skills	1.94	.25	108	1.57	.59	79
Extensive student role-play	1.58	.60	108	.90	.73	79
Adult feedback when students use skills	1.74	.50	108	1.37	.62	79
Learning objectives communicated to parents	1.61	.54	108	1.04	.69	78
Records kept of delivery of instruction	1.67	.60	108	.78	.73	79

Treatment Effect on Character Education Implementation						
Composites	Coefficient	*SE*	*df*	*P*	*ICC*	*ES*[c]
Materials	1.41	.14	10	.00	.39	1.85
Activities	0.82	.11	10	.00	.53	1.28

Note. The top panel shows descriptive statistics for treatment and control schools. The bottom panel shows tests of significant differences between materials used and activities delivered in treatment and control schools.

[a] *N* for the Materials composite is based on 1st and 2nd grade teachers; *N* for the Activities composite is based on 1st, 2nd, 3rd, 4th, and 5th grade teachers.

[b] 0 to 2 scaling applies to remaining 11 items.

[c] Effect size is calculated by dividing the coefficient by the standard deviation of the control group.

Appendix B

Table B

Demographic Characteristics of Baseline and Follow-up Sample

Sample		Percentage of Sample		
	N	Male	Academically Advantaged	Free or Reduced Meals
Baseline				
Treatment	985	50	91	11
Control	825	54	85	11
Total	1810	52	89	11
Teacher follow-up				
Treatment	574	49	94	8
Control	427	55	90	11
Total	1001	52	92	9
Parent follow-up				
Treatment	481	49	95	7
Control	351	54	92	9
Total	832	51	94	7

Appendix C

Table C1

Component Matrix for Rotated Two-Factor Solution for Teacher Ratings

Item	Component 1	Component 2
Solves problems with peers through compromise or discussion	0.81	-0.21
Tries a new approach to a problem when first approach is not working	0.79	-0.20
Is able to see things from other children's perspectives	0.77	-0.29
Expresses concern for others	0.75	-0.20
Articulates different ways to solve a problem	0.75	-0.03
Takes other peoples feelings into account before acting	0.73	-0.42
Removes him or herself from potential problem situations	0.72	-0.32
Helps others	0.71	-0.26
Resists peer pressure when appropriate	0.70	-0.34
Asks adult for help or advice about ways to resolve difficult situations	0.69	-0.01
Understands the likely consequences of his or her own actions	0.67	-0.36
Takes responsibility for own actions (for example, apologizes)	0.67	-0.46
Takes time to calm down when dealing with problem situations	0.66	-0.31
Lets others know how he/she feels about situations	0.65	0.17
If provoked by peers, shows self-control	0.59	-0.45
Shows respect for others	0.59	-0.56
If angered, expresses anger without being aggressive or destructive	0.43	-0.36
Teases, insults, provokes, or threatens others	-0.19	0.73
If upset, responds with verbal aggression (swearing, calling names)	-0.13	0.72
Responds with physical aggression to problems with peers	-0.17	0.70
Hits, kicks at, or jumps on other children	-0.12	0.67
Is impulsive in interacting with peers	-0.33	0.64
Acts in ways that annoy or bother others	-0.43	0.64
Gossips or spreads rumors	-0.12	0.63
Takes or steals things that belong to others	-0.01	0.62
Shows defiance in interactions with adults	-0.21	0.62
Tells lies or cheats	-0.19	0.61
Acts without thinking	-0.47	0.56

Appendix C

Table C2

Component Matrix for Rotated Two-Factor Solution for Parent Ratings

Item	Component 1	Component 2
Takes other people's feelings into account before acting	0.68	-0.33
Solves problems with peers through compromise or discussion	0.67	-0.18
Is able to see things from other children's perspectives	0.65	-0.20
Tries a new approach to a problem when first approach is not working	0.65	-0.05
Removes him or herself from potential problem situations	0.61	-0.23
Understands the likely consequences of his or her own actions	0.60	-0.17
Expresses concern for others	0.59	-0.18
Shows respect for others	0.59	-0.41
Helps others	0.58	-0.23
Resists peer pressure when appropriate	0.58	-0.31
Articulates different ways to solve a problem	0.57	-0.03
Takes responsibility for own actions (for example, apologizes)	0.56	-0.30
Asks adult for help or advice about ways to resolve difficult situations	0.55	0.04
Takes time to calm down when dealing with problem situations	0.55	-0.18
If provoked by peers, shows self-control	0.50	-0.42
Lets others know how he/she feels about situations	0.49	0.08
If angered, expresses anger without being aggressive or destructive	0.37	-0.39
Hits, kicks at, or jumps on other children	-0.13	0.68
Teases, insults, provokes, or threatens others	-0.17	0.62
Responds with physical aggression to problems with peers	-0.21	0.60
If upset, responds with verbal aggression (swearing, calling names)	-0.20	0.60
Shows defiance in interactions with adults	-0.21	0.58
Tells lies or cheats	-0.16	0.57
Acts in ways that annoy or bother others	-0.25	0.57
Takes or steals things that belong to others	-0.04	0.56
Gossips or spreads rumors	0.09	0.50
Is impulsive in interacting with peers	-0.12	0.47
Acts without thinking	-0.32	0.45

Appendix D

Table D

Frequency and Percentage of Students Rated in the Problem Range by Parents

School Number	Number of Problem Students		Percentage of Problem Students	
	Fall 2004	Spring 2007	Fall 2004	Spring 2007
		Control		
1	9	2	11	4
5	12	10	11	11
7	15	10	14	11
9	10	3	15	6
10	5	4	9	11
11	6	5	17	18
Total	57	34	13	10
		Treatment		
2	6	11	6	11
3	10	10	9	9
4	12	12	21	30
6	10	10	7	8
8	10	5	16	9
12	13	9	17	16
Total	61	57	11	12

References

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

Bloom, H.S., Richburg-Hayes, L., & Black, A.R. (2007). Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions. *Educational Evaluation and Policy Analysis, 29*, 30 – 59.

Catalano, C., Berglund, M., Ryan, J., Lonczak, H., & Hawkins, D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention and Treatment, 5*, 98 – 124.

Chung, H. & Elias, M. (1996). Patterns of adolescent involvement in problem behaviors: Relationships to self-efficacy, social competence, and life events. *American Journal of Community Psychology, 24*, 771-784.

Committee for Children (2009). *About us.* Retrieved May 8, 2009, from http://www.cfchildren.org/about/.

Committee for Children (2002). *Second Step Administrator's Guide: Preschool/Kindergarten – Grade 5*. Seattle, WA: Committee for Children.

Consortium on the School-based Promotion of Social Competence (1994). The school-based promotion of social competence: Theory, research, practice, and policy. In R. Haggerty, L. Sherrod, N. Garmezy, & M. Rutter (Eds.), *Stress, Risk, and Resilience in Children and Adolescents* (pp. 268 – 316). New York, NY: Cambridge.

Cooke, M.B, Ford, J., Levine, J., Bourke, C., Newell, L., & Lapidus, G. (2007). The effects of city-wide implementation of Second Step on elementary school student's prosocial and aggressive behaviors. *The Journal of Primary Prevention, 28,* 93-115.

De Los Reyes, A. & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483 – 500.

Deemer, S. & Minke, K. (1999). An investigation of the factor structure of the teacher efficacy scale. *Journal of Educational Research, 99,* 3 – 10.

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.

Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Second Step: Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders, 8,* 102-112.

Frey, K. S., Nolen, S.B., Edstrom, L.V.S., & Hirschstein, M. K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Applied Developmental Psychology, 26,* 171-200.

Gottfredson, D. (2001). *Schools and delinquency*. Cambridge, UK: Cambridge University Press.

Gottfredson, G. D., Gottfredson, D. C., Czeh, E. R., Cantor, D., Crosse, S., & Hantman, I. (2000). National study of delinquency prevention in schools. Ellicott City, MD: Gottfredson Associates (ERIC No. ED 459 409).

Gottfredson, D. C., Gottfredson, G. D., & Skroban, S. (1998). Can prevention work where it is needed most? *Evaluation Review, 22,* 315 – 340.

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.

Gottfredson, G., Nebbergall, A., Nese, J., Harak, E., & Shaw, F. (2008, March). *An Experimental Evaluation of Three-Years of Implementation of an Elementary School Social Competency Instructional Program*. Symposium conducted at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Greenberg, M. T., Coie, J. D, Lengua, L. J., & Pinderhughes, E. E. (1999). Predicting developmental outcomes at school entry using a multiple-risk model: Four American communities. *Developmental Psychology, 35,* 403 – 417.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnick, H. et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58,* 466 – 474.

Gresham, F. M. & Elliott, S. N. (1990). *The Social Skills Rating System*. Circle Pines, MN: American Guidance Service.

Grossman, D.C., Neckerman, H.J., Koepsell, T.D., Liu, P.Y., Asher, K.N., Beland, K., et al. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association, 277,* 1605-1611.

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., et al. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine, 33(2S),* 114-129.

Herrenkohl, R., Maguin, E., Hill, K., Hawkins, J. D., Abbott, R., & Catalano, R. (2000). Developmental risk factors for youth violence. *Journal of Adolescent Health, 26,* 176–186.

Holsen, I., Smith, B., & Frey, K. (2008). Outcomes of the social competence program Second Step in Norwegian elementary schools. *School Psychology International, 29,* 71 – 88.

Jessor, R. Van Den Bos, J., Vanderryn, J., Costa, F. M., & Turbin, M. (1995). Protective factors in adolescent problem behavior: Moderator effects and developmental change. *Developmental Psychology, 31,* 923 – 933.

Jessor, R., Turbin, M., & Costa, F. M. (1998). Risk and protection in successful outcomes among disadvantaged adolescents. *Applied Developmental Science, 2,* 194 – 208.

Kellam, S., Branch, J., Agrawal, K., & Ensminger, M. (1975). *Mental health and going to school*. Chicago, IL: University of Chicago Press.

Kellam, S., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results from a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry, 35,* 359 – 382.

Leffert, N., Benson, P. L., Scales, P. C., Sharma, A. R., Drake, D. R., & Blyth, D. A. (1998). Developmental assets: Measurement and Prediction of risk behaviors among adolescents. *Applied Developmental Science, 2,* 209-230.

Linares, L. O., Rosbruch, N., Stern, M., Edwards, M., Walker, G., Abikoff, H., & Alvir, J. (2005). Developing cognitive-social-emotional competencies to enhance academic learning. *Psychology in the Schools, 42*, 405 – 417.

Lipsey, M. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological Bulletin, 94*, 152 – 165.

Magazine, S., Williams, L., Williams, M. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's Affective and Continuance Commitment Scales. *Educational and Psychological Measurement, 56,* 241-250.

McMahon, S. D. & Washburn, J. J. (2003). Violence prevention: An evaluation of program effects with urban African American students. *Journal of Primary Prevention, 24*, 43 – 62.

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. Committee on Prevention of Mental Disorders, Institute of Medicine. *Reducing Risks for Mental Disorders: Frontiers for Prevention Intervention Research*. Washington, DC: National Academy Press; 1994.

Nation, M., Peabody, C., Crusto, C., Wandersman, A., Kumpfer, K., Seybolt, D.,

Morrisey-Kane, E., & Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist, 58,* 449 – 456.

Nebbergall, A. J. (2007). *Assessment of social competency and problem behavior: The psychometric properties of a social competency rating form.* Unpublished master's thesis, University of Maryland, College Park.

Nock, M. K., Kazdin, A. E., Hiripi, E., Kessler, R. C. (2007). Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey Replication.Preview. *Journal of Child Psychology and Psychiatry, 48,* 703-713.

Pellegrini, D. S. (1990). Psychosocial risk and protective factors in childhood. *Developmental and Behavioral Pediatrics, 11,* 201 – 209.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reiss, D. & Price, R. H. (1996). National research agenda for prevention research. *American Psychologist, 51,* 1109 – 1115.

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

Rubin, D. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91,* 473 – 489.

Rutter, M. (1979). Protective factors in children's responses to stress and disadvantage. In M. Kent and J. Rolf (Eds). *Primary Prevention of Psychopathology, Vol 3.* Hanover, NH, University Press of England.

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal*

of Orthopsychiatry, 57, 316-331.

Scales, P., Benson, P., Leffert, N., & Blyth, D. (2000). Contribution of developmental assets to the prediction of thriving among adolescents. *Applied Developmental Science, 4*, 27 – 46.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin.

Spivack, G., Marcus, J., & Swift, M. (1986). Early classroom behaviors and later misconduct. *Developmental Psychology, 22*, 124 – 131.

Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8*, 94–101.

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D., & Wikstrom, P. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal of Consulting and Clinical Psychology, 70*, 111 – 123.

Strein, B. & Koehler, J. (2008). Best practices in developing prevention strategies for school psychology practice. Chapter in Thomas, A. & Grimes, J. Best Practices V. Bethesda, MD: National Association of School Psychologists.

Taub, J. (2001). Evaluation of the Second Step violence prevention program at a rural elementary school. *School Psychology Review, 31*, 186 – 200.

Vera, E. & Gaubatz, M. D. (2002). Promoting social competencies in school-aged children. In C. L. Juntunen & D. R. Atkinson (Eds.), *Counseling Across the Lifespan* (pp. 57-70). Thousand Oaks, CA: Sage Publications.

Webster-Stratton, C. & Lindsay, D. W. (2002). Social competence and conduct problems in young children: Issues in assessment. *Journal of Clinical Child Psychology, 28,* 25-55.

Wilson, S. J. & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33,* S130 – S143.