2014

**Moderating Variables in the Treatment Effects of Second Step**

Gregory Eugene Moy  
*Loyola University Chicago*

Follow this and additional works at: [https://ecommons.luc.edu/luc_diss](https://ecommons.luc.edu/luc_diss)

Part of the *Educational Psychology Commons*

**Recommended Citation**

*Dissertations*. 1293.  
[https://ecommons.luc.edu/luc_diss/1293](https://ecommons.luc.edu/luc_diss/1293)

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact *ecommons@luc.edu*.

This work is licensed under a *Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License*.  
Copyright © 2014 Gregory Eugene Moy
LOYOLA UNIVERSITY CHICAGO

MODERATING VARIABLES IN THE TREATMENT EFFECTS OF SECOND STEP

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN SCHOOL PSYCHOLOGY

BY
GREGORY EUGENE MOY
CHICAGO, ILLINOIS
AUGUST 2014
Copyright by Gregory Eugene Moy, 2014
All rights reserved.
ACKNOWLEDGEMENTS

I would like to thank several organizations and individuals who have supported my graduate education over the past five years. I would like to thank the Loyola University faculty members who recognized my potential for scholarship through the application materials I submitted to the School Psychology program in 2009. I have received so much support from my faculty members since that time, and I am very grateful for that. Early on, Dr. Pam Fenning advocated on my behalf to secure funding for my education through the Illinois Board of Higher Education. I wish to extend my gratitude to Marcela Gallegos for coordinating all things related to my fellowship. Dr. Gina Coffee deserves thanks for providing me with an audience for the tangential trivia and corny jokes that I am inclined to share among other thoughts and opinions. Dr. David Shriberg, Dr. Martie Wynne, and Dr. Lynne Golomb deserve thanks for taking the time for extended conversations about matters related and unrelated to school psychology. I continue to appreciate your words of wisdom and advice. Thanks to Dr. Rosario Pesce, Dr. Dennis Simon, and Mr. Anthony Adamowski for leading by example and for fielding tough questions about translating theory to practice. I would like to thank Dr. Adam Kennedy and Dr. Terri Pigott for convincing me of the value of enthusiasm for research methods and for furthering my understanding of quantitative meta-analysis and research synthesis. Thanks to Josh Polanin for sharing in this.
enthusiasm and for steering me away from methodological pitfalls. Thanks to Alissa Briggs for collaborating on so many projects and to Brian Trainor for being reliable.

Thanks also to Tracy Ruppman for guiding me through the early stages of my literature search. I would also like to acknowledge the friends I have made through school. You are some of the kindest, most dedicated people I have met, and I am proud to be in your company working in the interest of children and the future.

I thank my dear friends who have known me for so long, who have stood with me through hard times, and who appreciate that we are all works in progress. I am grateful that we have shared so many experiences with one another, and I am looking forward to so many more as our families grow. I thank my parents for giving all the support they provide. I have learned from you in innumerable ways, and I am enriched by the experiences you have provided me. Thank you for allowing me to figure it out. I give thanks to my lovely bride whose kindness, tenderness, and patience restores me, fortifies me, and inspires me to carry on. You are steadfast when I am mercurial and balancing in so many ways. Emmitt, you are and always have been my inspiration. Daddy loves you.
## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .......................................................... iii

LIST OF TABLES ........................................................................ vii

LIST OF FIGURES ..................................................................... viii

ABSTRACT ................................................................................ x

CHAPTER I: INTRODUCTION ...................................................... 1

CHAPTER II: LITERATURE .......................................................... 7
  Youth Violence ......................................................................... 7
  Theoretical Considerations ..................................................... 11
    Applications of the Social Ecological Model to Understanding and Preventing Youth Violence ....................................................... 13
  School-based Prevention of Youth Violence ................................. 15
  Systematic Reviews of Social Emotional Learning (SEL) and Youth Violence
    Prevention Programs ............................................................. 19
    Evidence from Evaluations of SEL and School-based Violence Prevention Programs ......................................................... 21
  Description of Second Step ...................................................... 30
    Program Goals ....................................................................... 31
    Foundations ........................................................................... 31
    Revision History of Second Step ............................................. 32
    Organization and Implementation of Curricula ............................ 34
  Measuring Outcomes of Second Step ........................................... 37
    Moderators and Variation in the Treatment Effects of Universal Prevention Programs ........................................................... 38

CHAPTER III: METHODS ............................................................ 43
  Research Questions and Hypotheses ........................................... 43
  Variables ................................................................................. 46
  Eligibility Criteria ..................................................................... 47
    Dealing with Multiple Reports on the Same Sample .................... 48
  Search and Retrieval of Studies ................................................ 48
    Selection of Studies ............................................................... 51
  Coder Training and Reliability .................................................. 52
  Coding and Data Collection ...................................................... 53
  Data Analyses .......................................................................... 54
    Effect Size ............................................................................. 54
  Computational Models .............................................................. 56
    Test of Effect Size Homogeneity ............................................. 58
  Moderator Analyses ................................................................. 59
    Geography .............................................................................. 59
LIST OF TABLES

Table 1. Examples of Primary Study Outcomes ................................................................. 47
Table 2. Synopsis of Coding Protocol ................................................................................. 54
Table 3. Sources of Retrieved Studies ................................................................................. 64
Table 4. Study Screening Reliability .................................................................................... 65
Table 5. Title and Abstract Screening Results ..................................................................... 65
Table 6. Full Text Screening Results .................................................................................... 66
Table 7. Studies and Participants by Design and Outcome .................................................... 69
Table 8. Main Effect Homogeneity Tests .............................................................................. 75
Table 9. Studies Included in Geography Moderator Analysis ............................................. 77
Table 10. Studies Included in Implementation Scale Moderator Analysis ............................ 80
Table 11. Studies Included in Training Moderator Analysis .............................................. 84
Table 12. Studies Included in Dependent Variable Reporter Moderator Analysis .............. 88
Table 13. Studies by Grade Level Package and Outcome ..................................................... 91
Table 14. Moderator Homogeneity Tests ............................................................................ 91
Table 15. Trim and Fill Adjustments on Studies Reporting Prosocial Outcomes ............... 93
Table 16. Trim and Fill Adjustments on Studies Reporting Antisocial Outcomes .............. 94
Table 17. Trim and Fill Adjustments on Studies Reporting Knowledge Outcomes .......... 96
LIST OF FIGURES

Figure 1. PRISMA Search Flow Diagram .................................................................................. 67

Figure 2. Forest Plot of Prosocial Outcomes for Single Group Repeated Measures Designs.................................................................................................................. 70

Figure 3. Forest Plot of Antisocial Outcomes for Single Group Repeated Measures Designs................................................................................................................................. 71

Figure 4. Forest Plot of Knowledge Outcomes for Single Group Repeated Measures Designs................................................................................................................................. 72

Figure 5. Forest Plot of Prosocial Outcomes for Independent Group Designs.................. 73

Figure 6. Forest Plot of Antisocial Outcomes for Independent Group Designs .............. 74

Figure 7. Forest Plot of Knowledge Outcomes for Independent Group Designs .......... 75

Figure 8. Forest Plot of Geography Moderator on Prosocial Outcomes.......................... 78

Figure 9. Forest Plot of Geography Moderator on Antisocial Outcomes....................... 79

Figure 10. Forest Plot of Implementation Scale Moderator on Prosocial Outcomes ...... 81

Figure 11. Forest Plot of Implementation Scale Moderator on Antisocial Outcomes..... 82

Figure 12. Forest Plot of Implementation Scale Moderator on Knowledge Outcomes.... 83

Figure 13. Forest Plot of Training Moderator on Prosocial Outcomes............................ 85

Figure 14. Forest Plot of Training Moderator on Antisocial Outcomes......................... 86

Figure 15. Forest Plot of Training Moderator on Knowledge Outcomes....................... 87

Figure 16. Forest Plot of Dependent Variable Reporter Moderator on Prosocial Outcomes ................................................................................................................................. 89

Figure 17. Forest Plot of Dependent Variable Reporter Moderator on Antisocial Outcomes................................................................................................................................. 90
Figure 18. Funnel Plot of Studies Reporting Prosocial Outcomes .................................. 92

Figure 19. Forest Plot of Prosocial Outcomes by Report Type ...................................... 93

Figure 20. Funnel Plot of Studies Reporting Antisocial Outcomes .................................. 94

Figure 21. Forest Plot of Antisocial Outcomes by Report Type ...................................... 94

Figure 22. Funnel Plot of Studies Reporting Knowledge Outcomes .................................. 95

Figure 23. Forest Plot of Knowledge Outcomes by Report Type ...................................... 96
ABSTRACT

Youth violence is a problem which has been studied through theoretical frameworks based on social ecological models of development. This has coincided with the development of programs intended to prevent and reduce the occurrence of youth violence and its negative effects. One approach to preventing youth violence has been to deliver social emotional learning programs to children and adolescents in school settings. Social emotional learning programs are designed to provide children with skills and strategies for increasing self awareness, social awareness, emotion management, problem solving, and thoughtful decision making. Increasing students’ familiarity and use of these prosocial strategies are thought to reduce violence and other forms of aggression and antisocial behaviors. One of the most prominent programs used in schools to increase social emotional competence and reduce violence is Second Step. Several research studies have examined outcomes for students who have participated in the Second Step program. The present paper describes a meta-analytic approach to synthesizing research findings from primary research on Second Step through a social ecological framework by testing the hypotheses that program effects on student outcomes are moderated by geography, scale of program implementation, training for teachers, grade level package, and dependent variable reporter. Overall, students participating in Second Step demonstrated increased knowledge related to the topic areas of the curriculum. Students demonstrated increased prosocial outcomes after participation
compared to pre-intervention levels of prosociality. Participation in the program was not associated with significant decreases in antisocial behavioral outcomes as rated by teachers or through student self-report. The effects of Second Step appear stable across geography, different levels of implementation scale, and teacher training.
CHAPTER I
INTRODUCTION

Growing societal concern about youth violence has led to efforts to understand its causes and consequences and efforts to identify ways to effectively reduce its occurrence. A clear understanding of youth violence is needed in order to develop effective programs for its prevention. The social ecological model of development is a useful theoretical framework for conceptualizing complex social issues such as youth violence. In fact, the social ecological model has been widely adopted by researchers seeking to gain an understanding of youth violence (Bender, Emslie, & Bender, 2010; Dahlberg & Krug, 2002). Over the past several decades, the synthesis of research on the root causes of antisocial behaviors such as youth violence has demonstrated that these problems are multiply determined, and that multiple pathways contribute to the perpetration of violence by young people.

Individual factors and environmental factors derived from families, peer groups, schools, communities, and cultural influences can contribute to the risk of youths perpetrating violence, and by the same token, individual factors and environmental factors can also serve as protection against the risk of perpetrating violence. In applying an ecological model to the development of youth violence, the contributions of individual and environmental influences to violence can be thought of as components of nested environmental contexts in which the development of violent youth behavior occurs. One
major task facing researchers is how to examine empirically the interactive environmental influences on the expression of violence. A related task is to examine the potential of these influences to prevent youth violence. A more detailed description of the implications of the theoretical framework proffered by the social ecological model on the issue of youth violence will be included in the literature review.

One of the key ideas supported by the social ecological model is the notion that youth violence is more heavily influenced by proximal processes in the microsystem of the social ecology – such as those involving family members, peers, teachers, and components of schooling – than by the processes of more distal macrosystem contexts (e.g., broad political and cultural systems). This theoretical stance has contributed to the development of school-based research and programs intended to curtail the negative outcomes associated with youth violence. A vast number of violence prevention programs have proliferated in the past few decades. A good deal of the variation among these efforts is attributable to the various program goals and logic models espoused. For example, some programs are intended to achieve their desired outcomes through cognitive-behavioral changes in students, others aim to teach social skills through modeling and reinforcement, others aim to impart social and emotional competencies that are meant to reduce and resolve conflict, and others attempt to integrate multiple approaches and perspectives (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002).

The efficacy and effectiveness of many of these programs have been evaluated, generating mixed results. Some of the persistent issues in evaluating violence prevention programs and social-emotional learning programs are related to challenges in measuring the nebulous constructs involved. Systematic reviews have attempted to provide some
clarity to the field of school-based youth violence prevention and social emotional learning by synthesizing research findings from primary evaluation studies. Because of the complex nature of the problem of youth violence, the ecological complexity of schools and the variety of strategies employed by different school-based programs, researchers are challenged with partialing out the respective contributions of programs, contexts, and the characteristics of study samples in determining the effectiveness of prevention efforts. The complex issues dealt with in this field of research is especially salient when examining universal primary prevention programs. Unlike other types of programs which are intended to target selected samples based on risk factors, by definition universal violence prevention programs and social-emotional learning programs target wide and heterogeneous groups. In addition, implementation contexts can vary widely. Although the term “universal” connotes a wide catchment, it is a relative term when it comes to the feasible scale of implementation. Variation in the implementation of universal programs adds to the complexity of evaluating this approach to reducing the occurrence of violence. Several systematic reviews of school-based violence prevention and social-emotional learning (SEL) programs will be discussed in detail in the literature review. Particular attention will be paid to the differences in how these reviews have framed the problem of violence and aspects in the respective designs of the reviews which may contribute to inconsistent findings.

One of the most prominent and heavily researched universal programs in the fields of youth violence prevention and social emotional learning is Second Step. According to its publisher, Committee for Children, this pre-K-8 manualized curriculum has been translated from English into several other languages, and it has been
implemented in thousands of schools around the world since 1986 (www.committeeforchildren.org). Many evaluations of Second Step have been conducted in a variety of implementation settings, including some in countries outside the United States. The varying scales of Second Step implementation ranging from the level of single classrooms to entire districts have been evaluated. These evaluations have included participants who have differed in age, in nationality, in socioeconomic status, and across other demographic categories. According to social ecological theory, these differences may be substantive, and they ought to be meaningful in the appraisal of the effectiveness of Second Step.

The ecological influences that moderate youth violence may moderate the effectiveness of youth violence prevention programs such as Second Step. Because Second Step is intended to be implemented as a universal primary prevention program, it is intended to be delivered to all members of a targeted group – such as a classroom or school – without applying screening procedures or other forms of selection for participation. According to social ecological theory, individuals within a broad target population may vary in their exposure to risk and protective factors (Coie et al, 1993). It would follow that the heterogeneity of a population across dimensions of risk ought to produce variation in the impact of universal prevention programs. The literature on youth violence prevention does support the idea that intervention effects may differ across subgroups, especially for universal interventions such as Second Step.

Second Step curricula are organized by grade level. No known evaluations of Second Step have compared outcomes of students exposed to Second Step at different grade levels. Additionally, no known evaluations have examined implementation scale
as it relates to outcomes of students exposed to Second Step. The aim of the present research is to describe a meta-analysis of outcome studies evaluating the Second Step program. Social-ecological theory, a preferred framework with which to examine violence prevention, informs the identification of potential moderator variables that can be tested using meta-analysis. Overall effect size of Second Step will be calculated for three dependent variable categories: antisocial behaviors, prosocial behaviors, and knowledge related to Second Step content. Within each of these three dependent variable clusters, five potential moderators of effect size will be tested. These theoretically derived moderator categories are: geography, implementation scale, training for teachers, grade level, and dependent variable reporter. A detailed rationale for selecting these potential moderators, as well as methodological moderators, such as study design, will be provided in the literature review.

Following the literature review, a description of the methods used is presented. The first section of this chapter includes research questions and the hypotheses tested through meta-analysis. The second section describes the strategies used to identify studies eligible for inclusion. The third part of the chapter describes data collection and the coding of relevant study information. The fourth part describes the data analyses. Finally, limitations of the selected methods are discussed.

By isolating Second Step as a single input variable in the complex social ecology of youth violence prevention, the present meta-analysis is intended to further the understanding of the respective contributions of this curriculum by grade level, and it is intended to illuminate the degree to which selected aspects of the social ecology, such as implementation scale, moderate the effects of universal SEL programs on violence
prevention. A description of the corpus of studies on Second Step will be included in the results section. The synthesized overall effect sizes of the Second Step program is described, as well as the moderated effects of the program with respect to the types of outcomes reported by primary studies. In addition, other potential sources of variability in program effect that are not based in social ecological theory will be reported, such those related to study design and the as potential existence of positive results reporting bias in the corpus of literature on Second Step.

Finally, a discussion of the study’s findings concludes the present paper. A discussion of the practical implications of the findings includes recommendations for the implementation of the Second Step program and recommendations for future evaluations of the program following its implementation. The findings of the present study are discussed with regards to its broader implications on the social ecological theory of violence prevention and the role of universal SEL programs in school-based violence prevention efforts.
CHAPTER II

LITERATURE REVIEW

Youth Violence

Youth violence is a major social problem and public health issue for many communities. The problems and negative outcomes associated with school violence and aggressive behavior in youth have been well documented. Serious violence is one of the most pernicious threats to the health, well-being, and constructive potential of youth (Hall, Simon, Mercy, Loeber, Farrington, & Lee, 2012). The range of negative outcomes varies amongst perpetrators, victims, and individuals otherwise exposed to violence; they include physical, psychosocial, and economic harms (Ferguson, San Miguel, & Hartley, 2009; Lepore & Kliewer, 2013; Ramirez, Wu, Kataoka, Wong, Yang, Peek-Asa & Stein, 2012). These outcomes are matters of concern for students, school staff, and society at large (Daniels & Bradley, 2012; Fisher & Kettl, 2003; Price, Telljohann, Dake, Marsico, & Zyla, 2002).

According to the Centers for Disease Control, violence is the second leading cause of death for American youth ages 15-24. In 2010, 4,828 young people were victims of homicide. The lives of countless family members, peers, and community members are affected by these tragic deaths. The estimated medical and work costs attributed to homicides and injuries related to assaults amount to approximately $16 billion each year (Centers for Disease Control and Prevention, 2010). In 2011, 32.8% of
respondents in a nationally-representative sample of adolescent youth reported being in a physical fight in the 12 months preceding the survey. Twelve percent of survey respondents reported being in a physical fight on school property over this time span (Centers for Disease Control and Prevention, 2011).

The term violence has been defined in multiple ways, but most research definitions involve the actual or implied use of physical force to induce harm (Farrell & Flannery, 2006). Some definitions are broad, such as the one offered by Elliot, Hamburg, and Williams (1998) that violence is “the threat or use of physical force with the intention of causing physical injury, damage, or intimidation of another person.” Tolan and Guerra (1994) differentiated four types of youth violence: situational violence, relationship or interpersonal violence, predatory violence, and psychopathological violence. These experts on adolescent violence identified these four types as “existing on a multidimensional continuum within a biopsychosocial model of cause (p. 4). The causal etiology they advance shows differences across several important dimensions, such as the proportion of the population likely to show each type, the synergy of risk factors, and the likely age of onset, thus supporting their distinction. Situational violence is catalyzed by contextual risk factors such as extremely hot ambient temperatures, weekends, drug influence, and times of social stress. Relationship violence arises from interpersonal disputes between persons with ongoing relationships, in particular among family members and friends. Predatory violence is perpetrated intentionally to obtain some gain, and psychopathological violence is rare, but it is often extreme and virulent and seemingly without situational provocation. An alternative classification of violence offered by Fagan and Wilkinson (1998) focuses on the goals and functions of violent
behavior. Whereas Fagan and Wilkinson would agree with Tolan and Guerra that there needs to be more attention to subtypes of violence, they add that functions of youth violence can be summarized by: achieving and maintaining status, acquiring material goods and power, embarking on acts of retributive justice, displaying defiance towards authority, or following scripted behavior in contexts where violence is normative (Fagan & Wilkinson, 1998, pp. 64-79). The distinctions in the ways these groups of scholars have described violence underscore the need for researchers to further understand the nature of violence in order to reduce its occurrence among youth.

The Centers for Disease Control and Prevention Expert Panel on Protective Factors for Youth Violence Perpetration was convened to clarify some of these definitional issues to help address gaps in understanding violence among youth. A report of this panel’s work adopts a definition supplied by the Department of Health and Human Services that characterizes violence as an act which involves “the intentional use of physical force or power, threatened or actual, against another person or against a group or community that results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment, or deprivation” (Dahlberg & Krug, 2002). This definition associates intent with the act of violence, outcome notwithstanding. It is important to arrive at a consistent definition of youth violence for a number of reasons. A consistent definition is necessary in order to monitor the incidence of youth violence, to examine trends over time, and to uniformly measure risk and protective factors for victimization and perpetration. A well-developed information base about factors related to the perpetration of violence by youth may help improve strategies intended to reduce its occurrence and its negative consequences.
Studies of the etiology of violence have been used to guide the development of strategies to reduce the impact of youth violence by identifying risk factors and protective factors for youth violence perpetration. Risk factors are elements that predict an increased probability of a person acting violently (Kraemer, Lowe, & Kupfer, 2005). Protective factors include “attributes, characteristics, or elements that decrease the likelihood that violence will be perpetrated” (Hall et al., 2012). Most of the literature on predicting youth violence has focused on risk factors (Hawkins, Herrenkohl, Farrington, Brewer, Catalano, & Harachi, 1998; Herrenkohl, Maguin, Hill, Hawkins, Abbott, & Catalano, 2000; Loeber & Dishion, 1983). Much of the literature on protective factors focuses on processes of overcoming risk exposure (Loeber, Farrington, Stouthamer-Loeber, & Raskin-White, 2008; Losel & Farrington, 2012; Rutter, 1979; Werner, 2005).

The concept of risk and protection is not confined to the study of the development of violence in youth, but it is well-established as the prevailing paradigm for this field. The social ecological model is the most commonly used theoretical frameworks to organize risk and protective factors impacting the development of youth violence. The Centers for Disease Control and Prevention (CDC) are explicit and overt in their application of the social ecological framework to understand violence prevention (Dahlberg & Krug, 2002). This theoretical framework continues to be supported by empirical research on the etiology of violence which suggests that it is multiply determined by individual and environmental influences. The following section provides an overview of the social ecological model and its application to the understanding of youth violence.
Theoretical Considerations

This section will focus on the key concepts of the social-ecological model and its implications on understanding youth violence and its prevention. In the past three decades, scholars have advanced society’s understanding of violence based on social learning theory (Bandura, 1977), and further attempts to understand this problem have been informed by ecological models of development (Bronfenbrenner, 1979; Bronfenbrenner, 1994; Dahlberg & Krug, 2002). These theoretical foundations emphasize interactions between individuals and social contexts to explain the development of changes in thought and behavior. The social-ecological model has been used as a framework to conceptualize, develop, and evaluate prevention approaches and programs across multiple disciplines including violence prevention programs (Bender, Emslie, & Bender, 2010; Cunningham & Henggeler, 2001; Dahlberg & Krug, 2002; Gregson, Foerster, Orr, Jones, Benedict, Clarke, Hersey, Lewis, & Zotz, 2001; Herman, Merrell, Reinke, & Tucker, 2004).

The social ecological model recognizes that personal characteristics – including those of a biological and genetic nature – and social environments influence individuals’ behavior. Based on Urie Bronfenbrenner’s (1994) ecological systems theory, the social ecological model shares many of its defining features. The general ecological model, which lies at the center of Bronfenbrenner’s work, is defined by two core propositions. According to Bronfenbrenner, Proposition 1 states that “human development takes place through processes of progressively more complex reciprocal interaction between an active evolving biopsychological human organism and the persons, objects, and symbols in its immediate environment” (p. 38). Those immediate and enduring interactions in the
immediate environment are referred to as proximal processes. Proposition 2 states that “the form, power, content, and direction of the proximal processes affecting development vary systematically” as a joint function of the characteristics of the developing person, the environment, and the nature of the developmental outcomes under consideration (p. 38).

Bronfenbrenner (1994, 2001) emphasizes that proximal processes, or the interactions between the individual and the immediate environment, are the primary engines of development. In order to better understand the role of proximal processes, one must consider the nested structures that comprise the ecological environment as conceived in Bronfenbrenner’s model. These progressively nested structures are termed microsystems (in which proximal processes operate to produce and sustain development), mesosystems, exosystems, and macrosystems. Examples of microsystems include the home and other key developmental settings such as classrooms and schools which directly contain the developing individual. Mesosystems comprise the linkages between two or more microsystems, such as the relationship between home and school. Exosystems “comprise linkages and processes that take place between two or more settings, at least one of which does not contain the developing person” (Bronfenbrenner, 1994, p. 40). The macrosystem consists of the overarching pattern of the aforementioned structures. The chronosystem describes the time dimension of the model across which changes to the individual and the contexts occur (Bronfenbrenner, 1994).

The social ecological model formally adopted by the CDC as a framework for understanding violence includes four levels of influences on behavior: the individual, relationships, community, and society. Individual factors include biological and personal
history factors such as age, education, income, and history of abuse which impact the likelihood of becoming a victim or perpetrator of violence. In the second level of the model, factors related to relationships with close social-circle peers, family members, and partners are seen as impacting the likelihood of experiencing violence as a victim or perpetrator. The community settings such as schools, workplaces, and neighborhoods in which social relationships take place may influence risk of violence. Finally, the societal level of the model describes broad factors such as economics, social and educational policies, and cultural norms which create tolerance or intolerance for violence. According to this model, violence is the result of complex interactions between and among these four levels of factors (Dahlberg & Krug, 2002).

Applications of the Social Ecological Model to Understanding and Preventing Youth Violence

Although the CDC has chosen to focus on risk factors in their social ecological model of violence, protective factors at each of these levels may also exist. Whereas violent and aggressive behaviors may be learned and maintained through the social ecology, the corollary is that more adaptive, pro-social, and non-violent behaviors can also be acquired and maintained through the influences of the social ecology (Ladd, 1981). Recent research on risk and protective factors related to youth violence has been organized according to components of the social ecology. A recent national longitudinal study of youth violence categorized risk and protective factors according to the social ecological components of individual, family, school, and peer/neighborhood factors (Bernat, Oakes, Pettingell, & Resnick, 2012).
The social ecological conceptualization of youth violence has also served as the theoretical basis for the development of violence prevention programs. A four-dimensional grid model conceptualizing the typology of prevention efforts (Farrell & Camou, 2005) is consistent with the nested contexts derived from the social ecological model. Farrell and Camou’s typology categorizes youth violence prevention programs according to developmental stage, participants’ level of risk, and the level of the social ecology at which the program is intended to be effective. The fourth categorical dimension describes the goals of the prevention efforts, including: reducing levels of specific types of violence, reducing problem behaviors more generally, promoting social competence, and promoting positive youth development.

Ecological systems theory and the social ecological model both imply that school factors can play vital roles in the prevention of youth violence. Some of the school factors involved in risk models of youth violence include: low academic performance, low school commitment, low educational aspirations, school transitions, and peer delinquency (Bernat et al., 2012; Herrenkohl et al., 2000). School programming can also serve to protect youth against the risk of violence by providing students with opportunities to acquire skills and attitudes related to social and emotional competence that may preclude the use of violence. The results from a recently published study provide empirical support for the role of social competence as a mediating factor in the reduction of externalizing behavior problems, including violence (Langeveld, Gundersen, & Svartdal, 2012). The following section discusses the concept of social emotional learning, which has emerged as a prominent school-based approach to preventing youth violence.
School-based Prevention of Youth Violence

Among the various approaches to addressing youth violence which include surveillance, deterrence, and psychosocial interventions (Wilson & Lipsey, 2005), efforts that embrace prevention have drawn much attention. Prevention interventions are designed to disrupt the development of a problem by reducing exposure to risk factors and strengthening protective factors (Coie et al., 1993). Over 75% of schools in a U.S. national sample reported using a prevention program to deal with fighting, bullying, verbal conflict, and disruptive behavior, and many schools in the sample reported using multiple programs and strategies (Gottfredson, Gottfredson, Czeh, Cantor, Crosse, & Hantman, 2000). Catalano et al. (2002) contend that the most effective approach to reducing violence and other problem behaviors is to promote the development of social, emotional, cognitive, and behavioral skills. Many have argued that preventing violence requires teachers and school support personnel to become involved in providing students with strategies to regulate emotions, to solve problems, and to deal with conflict (Lopes & Salovey, 2004; Zins, Bloodworth, Weissberg, & Walberg, 2004).

Research-based approaches to violence prevention are necessary to ameliorate the negative outcomes for students, teachers, and their communities at large. Over the past few decades, an emphasis on social and emotional learning (SEL) has been frequently offered up as a solution for preventing violence and negative behavior in schools. School-based SEL programs have met considerable commercial popularity and success, yet a primary problem with determining research-based approaches to social and emotional learning as a viable violence prevention strategy is establishing a definition. A clear operational definition of SEL can provide educators, administrators, researchers,
and policymakers with a foundation on which assessment and evaluation can be based. In turn, the merits of efforts to enhance SEL in students as a form of violence prevention can be objectively judged.

The Collaborative for Academic, Social and Emotional Learning (CASEL) is widely regarded as the foremost authority on social and emotional learning. For nearly 20 years, CASEL has identified itself as an organization that “works to advance the science and evidence-based practice of social and emotional learning” (www.casel.org, n.d.). One of the most prominent definitions of social and emotional learning, as advanced by CASEL, is “the process of acquiring the skills to recognize and manage emotions, develop caring and concern for others, establish positive relationships, make responsible decisions, and handle challenging situations effectively” (CASEL, 2005, p. i). CASEL has characterized SEL as being comprised of five groups of inter-related competencies: self-awareness, social awareness, self-management, relationship skills, and responsible decision making (CASEL, 2003, p. 5). The definitions of these competencies, according to CASEL, are:

- **Self-awareness**: accurately assessing one’s feelings, interests, values, and strengths; maintaining a well-grounded sense of self-confidence.
- **Self-management**: regulating one’s emotions to handle stress, controlling impulses, persevering in addressing challenges; expressing emotions appropriately; and setting and monitoring progress toward personal and academic goals.
- **Social awareness**: being able to take the perspective of and empathize with others; recognizing and appreciating individual and group similarities and differences; and recognizing and making best use of family, school, and community resources.
- **Relationship skills**: establishing and maintaining healthy and rewarding relationships based on cooperation; resisting inappropriate social pressure; preventing, managing, and resolving interpersonal conflict; and seeking help when needed.
- **Responsible decision making**: making decisions based on consideration of ethical standards, safety concerns, appropriate social norms, respect for others, and likely consequences of various actions; applying decision-making skills to academic and social situations; and contributing to the well-being of one’s school and community. (Payton, Weissberg, Durlak, Dymnicki, Taylor, Schellinger, & Pacham, 2008, p. 6)

Based on this definition of SEL, one may be tempted to conclude that the measurement of these five competencies would accurately represent a measurement of SEL. However, empirical evidence in support of this characterization is limited. A factor analysis of the five competencies proposed by CASEL was conducted in the development of the Social-Emotional Learning Scale (SELS) (Coryn, Spybrook, Evergreen, & Blinkiewicz, 2009). Coryn and colleagues collapsed the competencies of social awareness and relationship skills together to create a Peer Relationships factor, and they collapsed the self-awareness and self-management competencies together as a Self-Regulation factor. The responsible decision-making competency was deemed homologous to the Task Articulation factor. Despite fitting the five CASEL competencies to a three-factor model, the researchers were unable to reject the null hypothesis of no difference between the model-implied covariances and the actual observed covariances. No other attempts to confirm the factor structure of SEL are known. Thus, no comprehensive measure of the process of social and emotional learning validated through statistical methods has emerged as the gold standard.

Instead, other proxy indicators have been used to establish the research evidence base for social and emotional learning. A variety of instruments have been selected by researchers to measure aspects of learning thought to be associated with SEL. In a systematic review of social function assessment tools for children and adolescents, 86
measures developed between 1988 and 2010 were identified (Crowe, Beauchamp, Catroppa, & Anderson, 2011). Some assessment tools identified were designed to broadly assess multiple areas of social function, for example, the Social Skills Rating Scale (Gresham & Elliott, 1990). Other tools are focused on a particular social skill, such as emotion recognition (Dyck, Ferguson, & Shocet, 2001). There is great variability in the research and evidentiary bases of the tools included in this review. The most highly-supported tools listed over 100 citations in their development literature, and others cited fewer than ten other papers as evidentiary or theoretical support. There is also great variability in the popularity of social function assessment tools. Popularity, as stated by Crowe and colleagues is influenced by accessibility, either through free download or commercial distribution (Crowe et al., 2011, p. 783). The variety of SEL-related measures provides options for educators interested in evaluating social competence as an approach to violence prevention, but making decisions about how to measure social competence can also be overwhelming.

Competence in the social arena involves several complex internal and external factors (Beauchamp & Anderson, 2010), and this complexity presents many challenges for measurement. The problems that exist due to the complex challenges associated with measuring SEL in students are echoed in attempts to measure the effectiveness of SEL programs. Researchers have attempted to measure the impact of SEL programs by selecting outcome measures which serve as proxy indicators of SEL, or they have chosen behavioral measures which are purported to be mediated by some aspect of SEL. If SEL is to be considered a safeguard against violence and aggression in schools, estimates of SEL program effectiveness should be made based on valid and reliable measures of SEL.
and on measures of violence or violence reduction. The related issues of defining SEL and measuring SEL are potential obstacles for educators and decision makers who are charged with executing policy objectives surrounding the promotion of SEL as a violence prevention strategy with students in schools. Given the absence of a clear gold standard assessment of SEL and the multitude of proxy measures used in research studies evaluating SEL, systematic reviews and meta-analyses are particularly well suited to drawing conclusions about the impacts of SEL programs and violence prevention programs on students.

**Systematic Reviews of Social Emotional Learning (SEL) and Youth Violence Prevention Programs**

The purpose of a systematic review “is to sum up the best available research on a specific question… by synthesizing the results of several studies” (www.campbellcollaboration.org, n.d.) using transparent procedures to find, analyze, and synthesize the results of relevant studies. The myriad outcome variables and intervention types used in primary research to approximate the acquisition and enhancement of SEL competencies may be organized and synthesized into broader, super-ordinate categories and evaluated through meta-analysis, which is the quantitative statistical analysis of several separate studies in order to test pooled data for statistical significance.

In addition to its application to the conceptual and practical understanding of social emotional learning as a potential violence prevention approach, meta-analysis may also be applied to the evaluation of violence prevention programs directly. Previous systematic reviews and meta-analyses on SEL programs and other violence prevention strategies have provided some insight as to how these interventions work at the universal,
secondary, and tertiary levels in school settings and in after-school programs. They may also yield evidence about generic intervention approaches or distinct program models. Meta-analysis can also shed light on the features of programs which contribute to effectiveness and the kinds of students who benefit most from violence prevention programs. The following section discusses the application of meta-analysis to the understanding of social emotional learning and the findings and limitations of systematic reviews focused on school-based programs to prevent youth violence.

In order to interpret the results of the following meta-analyses meaningfully, it is important to briefly explain some of the technical points on the methodology. Meta-analysis is the use of statistical methods to combine results of individual studies which examined similar treatments or interventions. One of the goals of meta-analysis is to improve estimates of treatment effects by synthesizing relevant information gathered from individual studies of the same intervention. This synthesized estimate of treatment effect is referred to as an effect size (ES). One of the most common effect sizes reported in meta-analyses is the standardized mean difference. The standardized mean difference is the difference between treatment group mean and control group mean divided by the pooled standard deviation. Standardized mean differences are usually presented as Cohen’s d or Hedge’s g. A standardized mean difference is generally considered statistically significant if zero is not included in its 95% confidence interval. Another widely used interpretation of effect size is to understand effect sizes represented by Cohen’s d of 0.2 as a “small” effect, around 0.5 a “medium” effect, and 0.8 to infinity a “large” effect. While the small-medium-large trichotomy is a widely used heuristic, arguments against this blunt interpretation of effect size have been presented. One of the
arguments against the generic application of the small-medium-large characterization of effect sizes is that it may be misleading, especially when considering the changes an effect size may represent when translated back into the initial units from which the effect size statistics are derived (Lipsey, Puzio, Yun, Hebert, Steinka-Fry, Cole, Roberts, Anthony, & Busick, 2012). Two other important statistics in interpreting the findings of meta-analyses are Cochran’s Q and I^2. Q is an indication of the heterogeneity of the studies included in the meta-analysis. It is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies. I^2 is also a description of the variation across studies included in a meta-analysis. It describes the percentage of variation across studies due to heterogeneity rather than chance.

**Evidence from Evaluations of SEL and School-based Violence Prevention Programs**

The breadth and inclusivity of CASEL’s definitions of SEL has permitted scholars to characterize numerous violence prevention, substance abuse prevention, health promotion, youth development, conflict resolution, self-esteem promotion, ethics-training, and social responsibility programs as falling under the SEL umbrella (CASEL, 2003, pp. 37-48). Hundreds of studies have been conducted on the impact this broad class of programs has had on students. A report released by CASEL in December 2008 summarizes the findings of hundreds of studies by means of three systematic scientific reviews (Payton et al., 2008). With regards to these three reviews of universal, selected, and after-school programs, the independent variable of interest is the implementation of an intervention that reviewers identified as a SEL program.

The same six categories of outcomes were analyzed and reported in the universal and indicated meta-analyses: (1) social and emotional skills, (2) attitudes toward self,
school, and others, (3) positive social behaviors, (4) conduct problems, (5) emotional distress, and (6) school performance. The presentation of the social and emotional skills outcome cluster was not presented in the results of after-school meta-analysis in this report. No explanation is offered for this omission. The first review presented findings from 180 studies which investigated the impact of SEL programs delivered as universal school-based interventions. These interventions were administered to a “general student body without any identified behavioral or emotional problems or difficulties.” The second review comprised 80 studies of school-based indicated programs. These programs were delivered to students who exhibited “early signs of behavioral or emotional problems.” The third review evaluated SEL programs delivered in after-school programs. These studies “primarily involved students without identified problems” (Payton et al., 2008, p. 5).

The universal review included 180 studies involving 277,977 students. The universal, classroom-based programming that these studies examined usually took the form of a specific curriculum or set of lessons intended to develop a variety of skills associated with social and emotional learning. Participants can be assumed to represent a general education population. Half (50%) of the outcomes summarized in this review are reflective of self-ratings by students; 20% by a teacher, 12% by independent observers, 4% by a parent, 2% by peers, and 11% were derived from school records. The results from the universal review presented in this report are summarized as mean effect sized (ES). The effect size variable (Cohen’s d or Hedge’s g) chosen for presentation was not identified. Therefore the abbreviation ‘ES’ was used to represent the reported effect size in this report. The number of studies from which the respective effect sizes were
calculated are represented by the abbreviation ‘k.’ Positive effects of SEL programs delivered as universal interventions were calculated for each of the six outcome categories: SEL skills (ES=0.60, k=56 studies), attitudes towards self and others (ES= 0.23, k=87 studies), positive social behavior (ES= 0.24, k= 84 studies), conduct problems (ES= 0.23, k=39 studies), emotional distress (ES = 0.23, k=39 studies), and academic performance (ES= 0.28, k=29 studies).

The indicated review included 80 studies involving 11,337 students. The studies included in the review focused on students who “showed signs of social, emotional, or behavioral problems, but had not been diagnosed with a mental disorder or need for special education” (Payton et al., 2008, p. 7). Approximately 38% of students participating in the studies displayed conduct problems, 23% displayed emotional distress, 10% displayed problems with peer relationships, and the remainder of studies focused on students who presented either a mixture of different problems or co-morbid problems. These studies relied on different parties to rate student outcomes. Thirty-eight percent (38%) of raters were the students themselves; 22% of ratings were by teachers; 11% by independent observers; 9% by peers; 7% by a parent; and 13% of ratings were derived from school records. Results presented in this review were promising. Positive effects were calculated for each of the six outcome categories: SEL skills (ES=0.77, k=11 studies), attitudes towards self and others (ES= 0.38, k=29 studies), positive social behavior (ES= 0.50, k= 38 studies), conduct problems (ES= 0.47, k=53 studies), emotional distress (ES = 0.50, k=35 studies), and academic performance (ES= 0.43, k=12 studies).
The after-school review included 57 studies involving 34,989 students. These programs were implemented outside of regular school hours, they were monitored by adults, and they had the goal of developing at least one personal or social skill. Positive effects were calculated for five out of the six outcome categories reported in the other two reviews: attitudes towards self and others (ES= 0.22, k=39 studies), positive social behavior (ES= 0.22, k= 33 studies), conduct problems (ES= 0.17, k=51 studies), emotional distress (ES = 0.91, k=5 studies), and academic performance (ES= 0.08, k=31 studies). Effect sizes for SEL skill acquisition or development were not reported, despite the centrality of that outcome category to the purpose of the review. In a paper authored by the lead researcher of these reviews, it is recommended that “researchers should report ESs for all outcomes regardless of their p-values” (Durlak, 2009, p. 918); thus the omission of these effect size statistics is puzzling.

Clearly, one of the strengths of meta-analysis is the flexibility it affords researchers in terms of the ability to cast eligibility criteria as widely or as narrowly as one chooses. The previously discussed meta-analyses have been instrumental in defining the field of social emotional learning through its broad inclusion of a variety of programs that may serve as protective factors in the social ecological framework of violence prevention. Other meta-analyses have focused on overlapping aspects of programs intended to prevent youth violence. A recent meta-review summarized findings across 37 meta-analyses and 15 systematic reviews of evaluations of behavioral and psychosocial approaches to preventing youth violence. A majority of the reviews included in the meta-review found moderate program effects. School-based program reviews generally had moderate to strong effects on youth violence-related outcomes (Matjasko, Vivolo-Kantor,
The authors acknowledge that there is a lack of independence between many of the reviews included. In other words, the same primary studies may have been included in more than one of the included reviews. This overlap is likely to impact the findings of the meta-review. Therefore, it may be useful to consider some of these reviews separately in an attempt to illustrate their differences in conceptual framing, eligibility criteria, and data reporting.

One meta-analysis focused on school violence prevention programs evaluated and published between the years of 1990 and 1999. Scheckner, Rollin, Kaiser-Ulrey, and Wagner (2002) calculated effect sizes for sixteen studies which used a control group and that implemented an experimental study design. This review did not calculate an overall weighted effect size for school-based programs evaluated in the included studies. Rather, the effects of each of the 16 included studies were discussed separately. Study samples ranged from 10 participants to 6,292 participants. Four of the studies demonstrated strong program effects, four studies contained intervention outcomes that produced moderate positive effects, and eight studies demonstrated small positive effects. Scheckner and colleagues were unable to support their hypothesis that a cognitive-behavioral theoretical orientation contributes significantly to program effect size based on the result that only one of the studies that demonstrated strong effects had a cognitive-behavioral basis. However, they did conclude that prevention efforts conducted at the elementary level were associated with strong effects. This conclusion was drawn based on the characteristics of three of the four studies that demonstrated strong effects.

A more methodologically rigorous meta-analysis was conducted by Jim Derzon (2006) which included 83 studies of 74 programs evaluated from 1950 to 2006. Findings
were pooled and synthesized by outcome type. Ten outcome categories related to antisocial and violent behavior were used: verbal aggression, disruptive aggression, aggressively inclined, carried a weapon at school, mixed crimes, fights, problem behavior, suspensions, mixed violence, and physical violence. Most of the evidence synthesized in this review came from samples ranging from 100 to 500 student participants. Based on this review, school-based interventions were typically successful in reducing antisocial behavior, but some notable exceptions exist. Thirty-four studies reporting findings on aggressive and disruptive behaviors yielded a small positive effect with a 5.5% reduction in these outcomes. Notably strong positive effects were reported for the 11 studies reporting reductions in criminal behavior outcomes by 20%. While indirect, suspensions as an indicator of violent and antisocial behavior were reduced by 12%. Studies reporting data on verbal aggression or carrying weapons in school did not demonstrate significant effects. Seven studies reporting data on problem behavior and two studies reporting violence and other antisocial behaviors demonstrated significant negative effects.

One of the most frequently cited meta-analyses on school-based interventions for aggressive and disruptive behavior is a review of psychosocial interventions conducted by Sandra Jo Wilson and Mark Lipsey (2007). This review updated a 2003 review with the inclusion of 77 additional outcome studies that met criteria of having been reported no earlier than 1950 and involved school-based programs from pre-kindergarten through 12th grade. Studies included reported at least one outcome variable representing aggressive, violent, or disruptive behavior, and studies used an experimental or quasi-experimental design (Wilson & Lipsey, 2007, p. S133). Data from 399 identified school-
based studies were extracted and analyzed. Eleven outcome categories were used in the synthesis: aggression, problem behavior, activity level/attention problems, anger/hostility/rebelliousness, social skills, social relations/adjustment, school performance, school participation, personal adjustment, internalizing problems, and knowledge/attitudes. All of the outcomes were positive and statistically significant with effect sizes ranging from 0.20 to 0.35. The overall weighted mean effect for universal programs on reducing aggressive behavior was 0.21; selected/indicated interventions produced a higher effect size of 0.29. The authors conclude that these effect sizes are not only statistically significant, but they also are likely to be of practical significance.

Whereas Wilson and Lipsey (2007) included both published and unpublished studies in their review, Park-Higgerson, Perumean-Chaney, Bartolucci, Grimley, and Singh (2008) restricted their literature search strategy to exclude dissertations, but they did not restrict by date of publication. Park-Higgerson and colleagues chose to include only studies which used an experimental design with randomized assignment to control or experimental groups. In total 36 studies were found eligible, and data synthesized from these studies were not as positive as those synthesized by Wilson and Lipsey (2007). Overall, no significant reductions in aggression or violence were found in intervention groups. Non-theory based interventions showed a larger effect than theory-based programs; selective programs showed larger effects than universal programs, and single-approach programs showed larger effects than multiple-approach programs. Student age also impacted effect; students in grades 4 and higher showed larger effects than students in grades 3 and lower.
A fourth meta-analysis on school-based violence prevention programs only included studies of universal interventions published prior to December 2004. Unlike the Park-Higgerson et al. (2008) review, which was restricted to published randomized controlled trials, the review authored by Robert Hahn and colleagues (Hahn et al., 2007) was limited to universal school-based programs. There is substantial overlap between this review and the subset of universal programs included in Wilson and Lipsey’s (2007) meta-analysis. Similar conceptual definitions were used, and similar outcome variables were assessed: self-reported or observed aggression or violence, measures of conduct disorder, measures of externalizing behavior, measures of acting out, measures of delinquency, and school records of suspensions or disciplinary referrals. One difference between the reviews is the timing of measured outcomes; Wilson and Lipsey used measures immediately following the intervention, whereas Hahn et al included follow-up data as well. Wilson and Lipsey included a broader array of literature sources, but Hahn et al. (2007) allowed a broader array of primary study methods. Hahn and colleagues suggest that their findings “are not greatly confounded” with the basic findings of Wilson and Lipsey’s (2007) review. Hahn et al. (2007) reported median effect by relative percent change in violent outcomes between intervention and control groups across several categories of program or study characteristics.

Three major conclusions have been made apparent by examining four separate systematic reviews and meta-analyses on school-based violence prevention programs. The first conclusion is that violence prevention programs tend to have an overall positive effect in reducing violence and aggression in participants. The second conclusion is that programs delivered as selected/indicated interventions tend to produce larger effects than
those delivered as universal programs. The third conclusion is that the specific inclusion criteria used in meta-analysis can be shaped to test a variety of hypotheses about violence prevention programs delivered in schools. Because several different programs and program types that were implemented in varied contexts with varying participants, challenges persist in the field’s ability to make meaningful generalizations about the effects of violence prevention programs on a variety of students based on previous meta-analyses.

In an effort to contribute to the understanding of school-based violence prevention, and more specifically, to the identification of program features that work better with different types of students, it would prove useful to isolate and synthesize the outcome research conducted on one of the direct factors derived from the social ecological model and examine other factors of the model as potential moderators of the direct factor. One of the most sensible direct factors to examine through meta-analysis has been program type. As seen in the previous meta-analyses, some focused on specific program types such as universal prevention (e.g., Hahn et al., 2007), and others included multiple types, but chose to examine program type (i.e., selected/indicated vs. universal) as a moderator of effect (e.g., Wilson & Lipsey, 2007). Even when limiting a synthesis to universal prevention programs, the heterogeneity of program features and implementation contexts create challenges in efforts to partial out and quantify the respective contributions of effects of program features, participant characteristics, and implementation contexts.

One potential solution to some of these persistent challenges is to identify a single social emotional learning program that has been implemented in several different
contexts with a variety of participants and to synthesize the outcome research on that one program. This would reduce the heterogeneity of program features to those that characterize that one singular program package. The direct and indirect forces by which a single social emotional learning program prevents violence can be examined without the introduction of confounds associated with including different programs in the synthesis. Conducting a single-program synthesis is similar to holding one independent variable constant and examining the effects of other relevant independent variables (e.g., participant characteristics, implementation contexts) on the final dependent outcomes. Such a study has the potential to not only provide meaningful data about the program of interest, but to provide additional information about the contributions of moderator variables to similar social emotional learning and violence prevention programs in general. As is true in all meta-analyses and research syntheses, a single-program meta-analysis relies on the authors of primary studies to report adequate information in order to conduct. The following section describes some of the research conducted on a program that is a strong candidate for a single-program synthesis.

**Description of Second Step**

Because resource allocation for programming needs often hinge on decisions regarding whether or not to purchase specific programs or packaged interventions, a synthesis of the research on a particular program may also have a high degree of practical utility. One of the most prominent and widely disseminated social and emotional learning programs targeting the prevention of violence is Second Step. This particular program has garnered recognition by several agencies and organizations for its quality and utility in schools seeking to prevent or remediate violence, aggression, and other
problem behaviors in students through the acquisition of social and emotional skills. For example, Second Step was granted an “Exemplary” rating by the U.S. Department of Education; it was given the distinction of a “Model” program by the Center for Substance Abuse Prevention; and it is regarded as a “SELect” program by the Collaborative for Academic, Social and Emotional Learning.

Program Goals

One of the primary assumptions held by the creators of Second Step is that young people are generally limited by social and emotional skill deficits, and that these skills can be taught in a school setting. Second Step is a violence prevention curriculum created with the intent to accomplish two main goals: the promotion of interpersonal and intrapersonal competencies, and reducing the development of social, emotional, and behavioral problems (Committee for Children, 1991, 1992a, 1992b, 1997). By providing direct instruction, modeling and observational learning opportunities Second Step aims to increase the interpersonal and intrapersonal skills of children exposed to the program. As a universal program, one of the goals is to cultivate a whole-school environment that addresses the social problems of children and supports the learning and use of positive social behaviors.

Foundations

Second Step is based on an eclectic blend of theories and conceptual frameworks, but it is primarily grounded in social learning theory. This foundation can be detected in the curricular emphasis on observation, self-reflection, and reinforced acquisition and performance of a desirable repertoire of behaviors (Bandura, 1986; Frey, Hirschstein, & Guzzo, 2000). Concepts and strategies derived from social information
processing (Dodge, Pettit, Mcclaskey, & Brown, 1986), cognitive-behavioral therapy (Kendall & Braswell, 1985), and Luria’s (1961) model of self-regulation through verbal mediation are integrated into a developmental sequence of social and emotional skill acquisition (Frey, Hirschstein, & Guzzo, 2000).

**Revision History of Second Step**

The first version of Second Step was originally published in 1986. According to a representative of Committee for Children, the publisher of Second Step, the program underwent its first revision in 1997 to update some of the visual aids and other materials included in the kit. Introductory units were added to each grade level curriculum at the request of clients, so that teachers and students could better prepare for upcoming lessons, and so that materials at each grade level could function as a standalone kit. These earliest versions organized preschool and kindergarten materials together, first through fifth grade materials together, and sixth through eighth grade materials together (E. Daggett, personal communication, June 24, 2013). The key thematic concepts of these first editions of Second Step were: Empathy, Impulse Control, and Anger Management. The focus on building empathic students is rooted in the belief that social and emotional competence is dependent on individuals’ ability to detect, understand, and respond sensitively to the feelings of others. The creators of *Second Step* argue that these requisite abilities are related to empathy, or the capacity to share the emotional state of another (Eisenberg, 1986; Frey, Hirschstein, & Guzzo, 2000). The second key concept is impulse control. The emphasis on impulse control is closely related to a careful and deliberate approach to problem solving. The lesson titles subsumed under the Impulse Control unit include: Identifying the problem, Choosing a Solution, Evaluating a
Solution, and Ignoring Distractions. The third key concept is anger management. *Second Step* teaches students how to recognize anger cues in their bodies and apply stress-reduction strategies to avert uncontrolled anger and angry behavior. The inclusion of this concept is intended to inhibit impulsive aggressive or antisocial behaviors through identifying anger triggers, calming down, and using positive self-talk.

In 2001, the third edition of Second Step was revised as “Second Step: A Violence Prevention Curriculum.” Some major aesthetic and conceptual changes accompanied this revision. New photos were added and reshot to better align with lesson content. The lesson cards were redesigned for teachers, and the number of lessons was reduced. Activities and exercises were updated for developmental appropriateness and to better hold student attention. More teacher modeling was also scripted into the lessons. The conceptual changes involved replacing the Impulse Control and Anger Management units with Emotion Management and Problem Solving. Emotion Management and Problem Solving are broader and more general ideas than Anger Management and Impulse Control. Furthermore, these reconceptualizations are framed with more neutral language which was viewed as a desirable program characteristic during the revision process (E. Daggett, personal communication, May 31, 2013).

The fourth edition of Second Step was revised in 2008 and released with three different names corresponding to the three different grade level packages. The preschool materials in the fourth edition of Second Step were separated from the kindergarten materials and organized under the subtitle, “Early Learning Program: Social-Emotional Skills for Early Learning.” The daily lessons provided in the Early Learning Program were made shorter to replace the longer weekly lessons in earlier versions. The
elementary grade materials are organized as, “K-5 Program: Skills for Social and Academic Success.” The materials for students in middle school are called, “Middle School Program: Student Success Through Prevention.” The middle school materials were revised at this time. In 2011, the materials for K-5 students were updated, and a Skills for Learning unit was added to preschool, kindergarten, and primary grade kits up to grade 3. Brain builder games were added to grades K-3, and online support, training, and supplemental resources were made available for all grades. Efforts to digitize the program included the production of multimedia lessons on DVD for grades 4-8 (E. Daggett, personal communication, May 31, 2013).

**Organization and Implementation of Curricula**

The curricula are designed to be presented in a classroom or group setting by teachers, school counselors, school social workers, school psychologists, and other youth service providers who have received program training. Common across all grade levels is the use of multimedia presentations, group discussions, modeling of relevant situations and behaviors, skill posters, and key concept visual aids. Depending on the grade level and edition of the program, curricula for preschool and elementary students consist of 15-28 lessons varying in length from 20-45 minutes intended to be presented once or twice per week.

Lessons are structured around large photo cards depicting age-appropriate social situations. On the backs of the cards, facilitators are provided with a suggested lesson script, key concepts, and objectives of the lessons. Teachers read these materials while asking students to engage in discussions and activities related to the situations depicted. These lessons are sometimes augmented by audio materials, such as songs for preschool
and kindergarten students which help normalize and provide vocabulary for experiencing feelings and emotions and strategies to cope with strong emotions. Audio and video materials for older children are provided to model salient behaviors, give instructions for coaching, and create opportunities to practice skills. Music videos about the key themes or concepts provide reminders and pneumonic devices to help students learn and remember strategies. Lesson videos offer simulations of realistic, age-appropriate situations in which key concepts and strategies may be employed. The lesson videos are broken up into short segments allowing for intermittent discussion throughout the presentation. The facilitation of group discussions can be guided through materials in the manual.

The earlier versions of the Second Step curriculum for middle school students is composed of 8 to 15 50-minute lessons per grade level organized into four units. The first unit describes violence as a societal problem, and it is centered on knowledge. The second unit encourages emotionality and empathy as ways to find common ground with others. It teaches children to avoid labeling and stereotyping. The third unit combines anger management training with interpersonal problem-solving with the goal of reducing impulsive and aggressive behaviors in adolescents. The fourth and final unit applies the skills learned in the previous units to five different situations. The desired behaviors are modeled in videos about making a complaint, dealing with peer pressure, resisting gang pressure, dealing with bullying, and diffusing a fight. More recent versions of the middle school kits focused on four similar and overlapping concepts: empathy and communication, bullying prevention, emotion management, and problem solving. A fifth
concept, substance abuse prevention, is included in the seventh and eighth grade materials.

Online interactive training for teachers and other school personnel is included with the classroom kits. Training can be completed by facilitators all at once, or in multiple sittings. In addition to the initial training modules, four additional staff training sessions are available for continuing training throughout the year. Teachers are encouraged to provide as many opportunities for discussion, role play, and real problem solving as possible. The manualized materials provided to teachers include a suggested lesson script that may be used to facilitate each lesson. In addition to the lesson script, key concepts are identified and organized for reference. The sequence of the lessons is pre-determined, and earlier lessons often serve as scaffolding for subsequent lessons. Lessons usually begin with a warm-up activity, and then the presentation of audio and visual media is used to stimulate group discussion. In addition to group discussion, some lessons include worksheet activities that can be completed individually or in small groups.

Behavioral skill training, modeling, and a combination of cueing, coaching and reinforcement are some of the eclectic teaching strategies used in Second Step to enhance the social and emotional learning of students. To underscore and complement classroom presentations, publishers of Second Step have also made “A Family Guide to Second Step” available to parents and caregivers of students. Each lesson is designed to include suggested complementary activities to be completed at home with caregivers or other family members to encourage generalization. Included in each grade level kit is a
developmentally appropriate summative knowledge assessment. The administration of this assessment is optional, but encouraged.

**Measuring Outcomes of Second Step**

Several evaluations and commentaries on Second Step have been published in peer-reviewed journals. Several more studies that focus on the implementation of Second Step have been conducted as dissertations and theses, or they have been conducted by governmental or non-governmental organizations. The degree to which researchers conducting primary outcome studies of Second Step have included the summative knowledge assessments that accompany the Second Step materials is unclear. Nevertheless, it may have a role in future research studies as one possible outcome measure. In the primary studies that have already been conducted, Second Step has been evaluated through the measurement of multiple dependent variables associated with antisociality, prosociality, and mental health. The outcomes related to antisociality have included: physical aggression, bullying, peer victimization, sexual violence, and other antisocial behaviors (Espelage, Low, Polanin, & Brown, 2013; Frey, Nolen, Van Schoiak Edstrom & Hirschstein, 2005; McMahon & Washburn, 2003). Prosocial outcomes measured have included: coping, cooperative behavior, conflict resolution, positive social behavior, problem solving, empathy, and social competence (Angelone, 2007; Cooke, Ford, Bourke, Newell, & Lapidus, 2007; Frey et al., 2005; Grossman et al., 1997; Holsen, Smith, & Frey, 2008; Taub, 2001). Knowledge acquired from the curriculum related to the contents of the program are measured in evaluations of Second Step using proprietary assessments included in the program materials (Edwards, Hunt, Meyers, Grogg, & Jarrett, 2005; Schick & Cierpka, 2005). The outcomes hypothesized to be impacted by Second
Step are broad and varied, yet they are all relevant to the discussion of SEL as a means to preventing school violence. Given this complex constellation of variables, decision makers may be interested in knowing what the overall effects of a universal SEL program such as Second Step have been, and what estimates of program effect they can reasonably expect.

**Moderators and Variation in the Treatment Effects of Universal Prevention Programs**

Universal prevention programs like Second Step are planned and implemented in the social contexts of youth (Farrell, Henry, & Bettencourt, 2013). Based on prior research in the fields of social emotional learning and violence prevention, we know that interventions don’t have the same effects on all people. There is variability in individuals’ responsiveness to interventions. This is especially true of universal interventions delivered to heterogeneous groups of individuals without screening for specific needs (Guerra, Boxer, & Cook, 2006; Langeveld, Gundersen, & Svartdal, 2012). Due to the heterogeneity of target populations exposed to universal prevention programs like Second Step, the question of effectiveness should not simply phrased as “does this work?” but as “for whom does this work, in what context, under what conditions, and according to whom?”

Violence and prevention of violence is multiply determined by risk factors and protective factors within the layers of the social ecological model and by the interactions of the components across layers. In this framework, a school-based universal prevention program delivered to youth in their classrooms is understood as operating within the microsystem of the social ecology. Other components of the social ecology of youth...
violence can provide suggestions for the identification of potential moderators of prevention program effectiveness. It is favorable to make a priori hypotheses about moderators that are derived within a theoretical model. Therefore, the following paragraphs outline potential moderators derived from the social ecological model: grade level, scale of implementation, geographical location, training of implementers, and dependent variable reporter.

According to the models offered both by Bronfenbrenner and the CDC, the individual is situated at the center of his/her own ecology. According to these models the characteristics of each individual ought to partially determine risk for violence. Because individual characteristics have an interactional relationship with ecological contests, they may also moderate the impact of interventions and programs delivered within the school microsystem to reduce violence. Because meta-analysis is limited to the data reported by authors of primary studies, it is reasonable to select potential individual-level moderators that are likely to be reported as participant demographics, such as age, grade level, and gender. Not only are these demographics useful from a pragmatic stance, but they are also well-researched variables in the theoretical and empirical understanding of social competence as a mediator of youth violence. As far as grade level is concerned, it has been determined that social competence and behavioral problems are more stable in older children than younger ones (Sorlie, Hagen, & Ogden, 2008). A synthesis on the variability of effects of Second Step by grade level may provide meaningful information regarding the relationship between social competence, age, and violence prevention. Gender is another potential candidate for selection as an individual-level moderator due to the likelihood of its being reported in primary studies, but Langeveld, Gundersen, and
Svartdal (2012) found no significant interactional effects of gender on social competence or behavior problems. Therefore, individual-level moderators of effect will be limited to student grade level in this study.

The scale to which a universal program is implemented may also play a significant role on the social ecology surrounding the expression of youth violence. According to the social ecological model, the implementation of a program like Second Step across an entire city or large school district has the potential to influence multiple levels of the exosystem in addition to the proximal processes of the microsystem. In turn, these exosystem influences ought to interact with the microsystem and the individual at the center of the model. Although Bronfenbrenner posits that the proximal processes of the microsystem are more influential in development than the distal processes of the exosystem and macrosystem, a program implemented on a large scale is likely to have a greater effect than the same program implemented in only one or a small handful of classrooms. According to Frey and Nolen’s transactional model of school-based prevention (2010), in addition to improvements in their own social and emotional competence, students trained with Second Step are also affected by the changes in their interactions with others trained in the program. This is another mechanism by which implementation scale may influence the effectiveness of Second Step.

To the degree that culture, government, economics, policies, and other macrosystem influences vary by geography, the physical locations in which Second Step is implemented may also moderate program effectiveness. It is known that Second Step has been translated into several languages and implemented in multiple countries. Outcome studies conducted in various geographical regions across the globe may further
our understanding of the role macrosystem influences impact violence prevention through this single program.

Another variable that may moderate the effect of Second Step is the type of training provided to teachers and other professionals implementing the program. The publisher of Second Step provides implementers with multiple training options that can be delivered remotely online or onsite at the school. School personnel may choose to use one of these authorized training modes, a combination of them, to conduct their own training of implementers, or to implement the program without training.

The effectiveness of a program like Second Step may vary across perspectives of different individuals positioned in the social ecology. That is, the individual at the center of his/her own social ecology may provide an indicator of the program’s effectiveness that differs from an indicator provided by someone else in that individual’s social ecology, such as a peer, a teacher, or a parent or caregiver. The potential variability in observed program effects based on these different perspectives may have significant implications on how to interpret program evaluations of Second Step and other universal SEL programs. Significant moderator effect by dependent variable reporter may have greater implications for the application of social ecological models of violence perpetration and prevention.

In addition to the substantive moderators that come out of applying a social-ecological model, other variables may also impact the observed effect of Second Step. Methodological variables, such as study design characteristics, including sampling techniques and selection of comparison groups have been studied in previous meta-analyses of violence prevention programs. Some of the meta-analyses which have
included these non-substantive moderators were intended to inform future evaluation research. Although the primary objectives of the present study are to inform decisions about implementing Second Step and to isolate program effects in order to examine the ecological moderators of effect on Second Step, some methodological moderators of effect are of interest for the sake of furthering the field of evaluation research in this arena. One potential methodological moderator may be the manner in which the outcome study was designed. A comparison of data generated from independent group design studies and single-group repeated measures designs may be informative.
CHAPTER III
METHODS

Given the fact that the study of youth violence and youth violence prevention has been a cooperative effort involving numerous scholars, researchers, and other stakeholders across several decades, trustworthy syntheses that accumulate past research findings are a necessary condition to orderly knowledge building (Cooper, 2010, p. 10). To date, no meta-analysis synthesizing the demonstrated effects of Second Step has been conducted to examine the potential moderating effects of implementation scale, geographical location, grade level, or dependent variable reporter. The following chapter provides a detailed description of the present meta-analysis of outcome studies of the Second Step program. This description is comprised of four parts. The first part consists of the research questions and hypotheses to be tested. The second part describes the study inclusion eligibility criteria used and the search strategy employed. The third part describes data collection and the coding of relevant study information. The fourth part describes the data analysis and moderator analyses. Finally, limitations of the selected methods are discussed.

Research Questions and Hypotheses

The present study was designed to address the following research questions and hypotheses:
1) What does existing evidence suggest regarding the effects of the Second Step program on the prosocial and antisocial behaviors of children in schools and on their knowledge and attitudes about violence prevention? In other words, what are the best estimates of the overall effect sizes of Second Step when outcomes are organized under categories reflecting prosocial behavior, antisocial behavior, and knowledge?

H₀: There is no difference between students who participated in the Second Step program and students who have not participated in the program on prosocial, antisocial, or knowledge outcomes.

H₁: Students who participated in the Second Step program experience increased prosocial outcomes, decreased antisocial outcomes, and increased knowledge related to violence prevention.

2) Do social ecological variables influence the effect of Second Step on students’ prosocial, antisocial, and knowledge outcomes? This query can be broken down into five separate questions:

a. Does implementation scale moderate the effects of Second Step?

H₀: Implementation scale has no statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, or knowledge outcomes.

H₁: Implementation scale has a statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, and knowledge outcomes.

b. Are there significant differences in the effects of Second Step across different geographical regions of implementation?

H₀: Geography has no statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, or knowledge outcomes.
H₀: Geography has a statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, and knowledge outcomes.

c. Does the presence and type of implementer training impact the effect of Second Step on student outcomes?
H₀: Implementer training has no statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, or knowledge outcomes.
H₁: Implementer training has a statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, and knowledge outcomes.

d. Do different grade level packages moderate the effects of Second Step?
H₀: The grade level package implemented has no statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, or knowledge outcomes.
H₁: The grade level package implemented has a statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, and knowledge outcomes.

e. Are there significant differences between student self-reports of outcomes and ratings provided by other individuals, such as teachers?
H₀: Dependent variable reporter has no statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, or knowledge outcomes.
H₁: Dependent variable reporter has a statistically significant moderating effect on the effects of Second Step on prosocial, antisocial, and knowledge outcomes.

3) Whereas the second research question probes into substantive moderators of effect, the third major research question poses the possibility that positive reporting bias exists among published studies of Second Step.
H₀: There is no difference between the effects reported in published and unpublished studies (dissertations) of Second Step on prosocial, antisocial, or knowledge outcomes.

Hₐ: Published studies of Second Step report more favorably on prosocial, antisocial, or knowledge outcomes than unpublished studies, such as dissertations.

Variables

Meta-analysis involves two categories of variables: descriptive variables which constitute the independent variables, and effect sizes, which constitute the dependent variables. The primary descriptive variables of interest in the present study consist of: (a) the geographical location of the study operationally defined as the continent on which the study was conducted, (b) the various grade-level packages of Second Step, which are Pre-Kindergarten, Elementary (K-5), and Middle School (6-8), (c) training of implementers, (d) the scale in which the program was implemented, and (e) the individual in the social ecology reporting the dependent variables. These independent variables will be described in greater detail in a following section about moderators.

Existing studies on Second Step have employed a variety of dependent variables to evaluate program effectiveness. In the present study, program effects will be analyzed according to dependent variables representing the three overarching constructs of: (a) prosociality, (b) antisociality, and (c) knowledge.

Prosocial outcomes include data gathered from observations or ratings of student behavior that are associated with the decrease or prevention of violence, such as empathy, altruism, cooperation, and perspective-taking. Antisocial outcomes include data gathered from observations or ratings of student behavior that are associated with the increase or
promotion of violence, such as various forms of aggression, fighting, or bullying. Knowledge outcomes include data collected from instruments testing knowledge or attitudes about violence or violence prevention, such as the proprietary unit assessments that are included in the Second Step kits.

Table 1. Examples of Primary Study Outcomes

<table>
<thead>
<tr>
<th>Examples of Prosocial Outcomes</th>
<th>Examples of Antisocial Outcomes</th>
<th>Knowledge Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive coping</td>
<td>Physical aggression</td>
<td>Second Step Knowledge and Skills Assessment</td>
</tr>
<tr>
<td>Cooperative behavior</td>
<td>Bullying</td>
<td></td>
</tr>
<tr>
<td>Impulse control</td>
<td>Peer victimization</td>
<td></td>
</tr>
<tr>
<td>Positive social behavior</td>
<td>Delinquency</td>
<td></td>
</tr>
</tbody>
</table>

**Eligibility Criteria**

One of the purposes for conducting a meta-analysis on a particular topic is to provide a meaningful synthesis of the existing research. As highlighted by the preceding review of other meta-analyses on school-based violence prevention programs, the inclusion criteria researchers establish when designing a systematic review are of central importance in terms of the findings that emerge from a synthesis, and these criteria also delimit a degree of the external validity of a systematic review.

Studies were selected for this meta-analysis based on a set of detailed criteria based on the overarching research questions. The eligibility criteria are summarized as follows:

1. The study must include an evaluation of the Second Step program implemented in a school setting.
2. The study assessed intervention effects for at least one outcome variable that represents either:
a. an antisocial outcome or a potential risk factor for violence,
b. a prosocial outcome or a potential protective factor against violence, or
c. a demonstration of knowledge related to the prevention of violence.

3. Unadjusted quantitative data must be reported for at least one outcome variable of interest.

4. The study used either of the following research designs:
   a. An experimental or quasi-experimental design that compared students exposed to the Second Step program with one or more comparison conditions on at least one qualifying outcome variable.
   b. A repeated measures design in which measures of at least one qualifying outcome variable were taken before and after intervention on the same participants.

5. The study report is available in an English language format.

**Dealing with Multiple Reports on the Same Sample**

Multiple papers evaluating the Second Step program on the same group of individuals are examined to determine which study reports information best suited for meta-analysis. One study is selected among these to avoid any single sample’s overrepresentation in the synthesis of program effects.

**Search and Retrieval of Studies**

An attempt was made to identify and retrieve the entire population of empirical evaluations of the Second Step program that meet the eligibility criteria specified above including both published and unpublished research reports. A comprehensive search for studies was conducted within the following 13 bibliographic databases, comprising the primary search strategy: Academic Search Complete, Child Development and Adolescent
Studies, Education Research Complete, Education Administration Abstracts, Educational Resources Information Center (ERIC), OmniFile Full Test Select, Professional Development Collection, PsycINFO, PsycArticles, Social Work Abstracts, Teacher Reference Center, Social Work Reference Center, and Dissertations and Theses (ProQuest). Secondary search strategies included hand-searches within targeted journals, searches within the reference lists of included studies and the reference lists of relevant meta-analyses.

The primary search strategy was comprised of multiple stages in multiple databases. In these 13 bibliographic databases, two sets of search terms were used. The first stage was comprised of a search of all the aforementioned databases for studies on Second Step comprised of the English program name, “Second Step” and additional search terms to improve the relevance of results due to the common usage of the phrase, “second step” in the English language when describing processes not necessarily related to the current topic. The terms “school program,” “school-based program,” “school intervention,” and “school-based intervention” were combined with the primary search term, “Second Step” in the following manner and comprised the first stage of the search:

[“Second Step” AND (“school program” OR “school-based program”)] OR
[“Second Step” AND (“school intervention” OR “school-based intervention”)].

The second stage was comprised of applying a set of search terms to the set of 13 databases for the different names given to the Second Step program’s translations. These names, which were identified on the publisher’s website, are: “Faustlos” (German), “Tulleriit” (Greenlandic), “Steg for Steg” (Norwegian), “Trin for Trin” (Danish), “StegVis” (Swedish), “Paso Adelante” (Spanish), “Antras Zingsnis” (Lithuanian),
“Askeleittain” (Finnish), “Srdce Na Dlani (Slovak), “Ikinci Adim” (Turkish), “Hengaw be Hengaw” (Kurdish), “Sekando Suteppu” (Japanese). The translated program names are combined as search terms in the following manner:

- “faustlos” OR “tulleriit” OR “steg for steg” OR “trin for trin” OR “stegvis” OR “paso adelante” OR “antras zingsnis” OR “askeleittain” OR “srdce na dlani” OR “ikinci adim” OR “hengaw be hengaw” OR “sekando suteppu”.

The sets of search terms described above were applied to the full text of studies, and results were date limited between 1984 and 2013 in all stages of the primary search. In addition to the date limit, one additional limit was applied to the Dissertations and Theses database to include only results from studies identified as dissertations.

After each stage of the primary search, results were exported from the online databases to separate folders created in the Mendeley Desktop reference and citation manager in RIS (Research Information Systems) format. Secondary search strategies were employed in addition to the primary online database searches. Reference lists of outcome studies found eligible for inclusion through the database search were examined for other outcome studies. Given the recent emergence of a number of research syntheses on programs addressing social skills, social competence, interpersonal competence, social emotional learning, violence prevention, and affective education, the references of closely related systematic reviews on school-based violence prevention (Derzon, 2006; Park-Higgerson, et al., 2008; Payton et al., 2008; Scheckner et al., 2002; Wilson & Lipsey, 2007) were also searched for potential outcome studies on Second Step eligible for inclusion in the present meta-analysis. The Second Step program’s publisher,
Committee for Children, cites a number of outcome studies on their website, and these were included in the search. Additionally, the following journals were targeted for a hand search: *Journal of School Violence, Journal of Primary Prevention, Prevention Science, Learning Community: An International Journal of Educational and Social Development, International Journal of Education Research, International Journal of Education, Journal of Scandinavian Studies in Criminology and Crime Prevention*.

Once candidate reports were collected, a coding team comprised of the primary researcher and a colleague screened the candidate reports for inclusion eligibility. A detailed description of this screening process is described in the following section.

**Selection of Studies**

The primary researcher and one other coder conducted all steps of the search strategy. The results of the online database searches were saved as potential candidates for inclusion in the Mendeley Desktop program. Coders examined saved report titles and abstracts to screen for eligibility. At this stage, only reports that clearly did not pertain to the Second Step program were excluded. An example of a report excluded at this phase would be one that contains the phrase “second step” as it refers to part of a process clearly not related to violence prevention, social emotional learning, or school-based programs or intervention, such as engineering or biochemistry. Another example of reports that were not included at this phase were evaluations of similar but distinctively different programs or curricula which were explicitly named in the title or abstract, such as the PATHS program or The Incredible Years program. The full text of studies included from the title and abstract screening were further screened to determine eligibility for inclusion in the meta-analysis. In the full text screening stage, coders
determined whether the report represents an outcome evaluation of the Second Step program. All other reports were excluded from the meta-analysis. Information from candidate reports that remained eligible for analysis were collected through a formal coding protocol.

**Coder Training and Reliability**

The second coder was trained on the screening procedures by the primary researcher. After the initial training and orientation to the screening procedures, 20% of the remaining candidate studies were coded independently by each of the coders to establish inter-coder reliability. An acceptable degree of reliability for the title and abstract screening is represented by a Cohen’s kappa of 0.81 or greater. After an acceptable level of reliability was achieved, all remaining reports identified as potentially eligible for inclusion were assigned to each of the two coders for title and abstract screening and full text screening. Results of the screening reliability check are described in Chapter IV. A detailed list of included studies is included in Appendix A.

Another coder was trained by the primary researcher on the coding of relevant study information once all included studies were identified through the screening procedures described above. A sample of eight candidate reports was coded together during the initial training. The sample of reports used in the training represent heterogeneous study characteristics, such as implementation scale, type of outcome variable measured, and study design. After this initial training, additional studies were coded by both coders to test the inter-coder reliability. Acceptable reliability was again represented by a Cohen’s kappa of 0.81 or greater. Once this level of reliability was
achieved, the remaining studies were coded. Results of the coding reliability are described in Chapter IV.

**Coding and Data Collection**

A coding protocol has been developed to systematically screen studies for eligibility and to encode two broad classes of information. The coding protocol encodes information about study characteristics, and it also encodes information about the empirical findings of each study (Lipsey & Wilson, 2001, p. 73). Study characteristics (e.g., source descriptors, methods, measures, samples, contexts, etc.) represent factors that may influence the magnitude and nature of the findings. Necessary effect size data was also coded. The development of the protocol for the present study has been informed by other related meta-analyses (Kennedy, 2010; Wilson, 2000), and the protocol is organized into six sections. Section A pertains primarily to bibliographic information and additional information about study design, which is relevant for the selection of studies. Information on the study context and the scale of implementation comprises Section B of the protocol. Section C encodes information about the implementation of the program, such as duration and training. Section D encodes information about the participants. Section E encodes characteristics about the outcome variables and the measurement instruments. Effect size data used in determining intervention effects was encoded in Section F. Each of these respective sections of the coding protocol contains information useful for conducting the meta-analysis. A complete coding protocol for the present study may be found in Appendix B. Table 2 outlines examples of information encoded.
Table 2. Synopsis of Coding Protocol

<table>
<thead>
<tr>
<th>Section</th>
<th>Examples of information encoded</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Bibliographic information and Screening</td>
<td>Study ID</td>
</tr>
<tr>
<td></td>
<td>Publication type</td>
</tr>
<tr>
<td></td>
<td>Publication source</td>
</tr>
<tr>
<td></td>
<td>Design type</td>
</tr>
<tr>
<td>B. Study and Implementation Contexts</td>
<td>Scale of implementation</td>
</tr>
<tr>
<td></td>
<td>Scale of evaluation</td>
</tr>
<tr>
<td>C. Intervention implementation</td>
<td>Implementation details</td>
</tr>
<tr>
<td></td>
<td>Implementation fidelity</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
</tr>
<tr>
<td>D. Participants</td>
<td>Number of participants</td>
</tr>
<tr>
<td></td>
<td>Age/Grade level of participants</td>
</tr>
<tr>
<td></td>
<td>Reported risk factors</td>
</tr>
<tr>
<td></td>
<td>Predominant ethnicity of participants</td>
</tr>
<tr>
<td>E. Dependent variable characteristics</td>
<td>Construct measured</td>
</tr>
<tr>
<td></td>
<td>Sources of outcome data</td>
</tr>
<tr>
<td></td>
<td>Reliability of outcome data</td>
</tr>
<tr>
<td>F. Effect Size Data</td>
<td>Quantitative values reported</td>
</tr>
<tr>
<td></td>
<td>Means</td>
</tr>
<tr>
<td></td>
<td>Standard deviations</td>
</tr>
<tr>
<td>G. Bias Analyses</td>
<td>Disclosed funding</td>
</tr>
<tr>
<td></td>
<td>Author affiliations with publisher</td>
</tr>
</tbody>
</table>

**Data Analyses**

**Effect Size**

Studies eligible for inclusion in the present meta-analysis present a comparison of Second Step treatment effects with some form of no-treatment or pre-treatment control. The appropriate effect size statistic used to synthesize these types of study-level comparisons is the standardized mean difference. The standardized mean difference represents the difference between treatment group and comparison group means on an outcome variable. Because different studies evaluating Second Step have used different instruments to assess outcomes, the raw mean differences would not be meaningful to
combine as they are. Dividing the differences in means by the standard deviation within
groups creates a standardized index that is comparable across studies. The sample
standardized mean difference, referred to as Cohen’s $d$, is represented by the equation:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}$$

Where $\bar{X}_1$ represents the mean of the treatment group, $\bar{X}_2$ represents the mean of the
comparison group, and $S_{within}$ represents the within-groups standard deviation, pooled
across groups,

$$S_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Where $n_1$ and $n_2$ are the sample sizes in the two groups, and $S_1$ and $S_2$ are the standard
deviations of the two groups. The variance in $d$ is given by

$$V_d = \frac{n_1 + n_2}{n_1n_2} + \frac{d^2}{2(n_1 + n_2)}$$

and the standard error of $d$ is the square root of $V_d$

$$SE_d = \sqrt{V_d}$$

It is known that $d$ can have a slight bias which tends to overestimate the absolute value of
the effect in small samples. The bias may be removed by the correction factor $J$ which
results in the unbiased estimate called Hedges’ $g$. An approximation of $J$ commonly used
is

$$J = 1 - \frac{3}{4df - 1}$$

in which $df$ represent the degrees of freedom used to estimate $S_{within}$. In order to obtain $g$,
Cohen’s $d$ is simply multiplied by the correction factor $J$
\[ g = J \times d \]

The variance of \( g \) is represented by the equation

\[ V_g = J^2 \times V_d \]

**Computational Models**

Because this is a synthesis of a single program, there is a temptation to apply a fixed effects model to the combination of effects from the full set of studies. A fixed effect model assumes that all the studies in the meta-analysis share a common true effect size. However, this temptation must be avoided for the following reason. There is no reason to assume that the included studies are “identical” in the sense that the true effect size is exactly the same in all the studies. Reasons that the true effect could vary from study to study include, but are not limited to the following: variation in the context of program implementation, variation in participant demographics and risk history, variation in age/grade of participants, variation in the quality of the instructional materials at each grade level, variation in the quality of the instructional materials between program revisions, variation in implementation fidelity, variation in the type of effect measurement, or unknown covariates.

A random effects model will be applied to the statistical analyses conducted in the present study. The random effects model carries with it the assumption that each observed effect size differs from the population mean by subject-level sampling error plus a value that represents other sources of variability assumed to be randomly distributed (Borenstein, Hedges, Higgins & Rothstein, 2009). The fact that Second Step is designed as a universal primary prevention program to be implemented with largely heterogeneous groups suggests that other sources of variability exist in addition to
subject-level sampling error. Thus, a more conservative, random effects model is preferred with respect to the likely heterogeneity of study samples. Under this model, it is assumed that a portion of the variability is systematic and can be statistically modeled, and another portion of the variability is random and cannot be modeled. This random variance component will be added to the estimation of variance associated with sampling error in calculations of program effect. Because it is assumed that this variance term reflects sources of variability not specific to any particular study sample, but rather to normally distributed variability in the population, the pooled estimate of variance will be applied across all study-level cases. Moderator analyses will also be conducted using a random effects model and a pooled variance term.

To compute the variance of a study under a random effects model, the between-studies variance is needed in addition to the within-study variance because the study’s total variance is the sum of these two values. The between-studies variance is represented by the parameter $\tau^2$ (tau-squared). A way to estimate this parameter is to compute

$$
\tau^2 = \frac{Q - df}{c}
$$

where

$$
Q = \sum_{i=1}^{k} W_i T_i^2 - \frac{\left(\sum_{i=1}^{k} W_i T_i\right)^2}{\sum_{i=1}^{k} W_i},
$$

$$
df = k - 1,
$$

Where $k$ is the number of studies, $W_i$ is the weight assigned to study $i$, $T_i$ is the observed effect of study $i$ and
Each study will be weighted by the inverse of its variance. Under a random effects model, the weight assigned to each study is represented as

\[ W_i = \frac{1}{V_{yi}^{\ast}} \]

where \( V_{yi} \) is the within-study variance for study \( I \) plus the between-studies variance \( T^2 \).

That is,

\[ V_{yi}^{\ast} = V_{yi} + T^2. \]

**Test of Effect Size Homogeneity**

In a meta-analysis, variation in outcomes between studies is known as heterogeneity. The \( Q \) statistic is the most common test of homogeneity, and it is calculated to examine the ratio of observed to expected dispersion in effect sizes. The \( Q \) statistic is distributed as a chi-square distribution with degrees of freedom of \( k - 1 \), where \( k \) is the number of studies analyzed in the meta-analysis. Therefore, if \( Q \) is higher than the critical value based on \( k - 1 \), the null hypothesis that the distribution is homogenous (variability is due to chance alone) will be rejected. In a random effects model, the \( Q \) statistic tests whether tau-squared is significantly different from zero. A rejection of the null hypothesis of homogeneity suggests that the variation in outcomes is due to more than sampling error alone (i.e., studies are heterogeneous). The \( Q \) statistic does not indicate the proportion of the observed variance due to heterogeneity (i.e., between-study variability). In order to determine the degree of variability in effect sizes, the \( I^2 \) index will be calculated using the following formula:
\[ I^2 = 100(Q - k)/Q \]

The interpretation of the \(I^2\) index was based on the guidelines provided by Higgins, Thompson, Deeks, and Altman (2003) in which an \(I^2\) less than 25\% means that the amount of variability among effect sizes is low, 50\% reflects a moderate amount of variability, and 75\% or greater denotes a high amount of variability.

**Moderator Analyses**

Moderator variables are hypothesized to affect the magnitude of the relationship between variables in this meta-analysis. Depending on the availability of information reported in the included studies, five potential moderators are proposed: geography, implementation scale, grade level package, implementer training, and dependent variable reporter. A fully random effects model is used within subgroups of each moderator variable. Although a separate variance may be calculated for each subgroup of a moderator, pooled variances will be used due to small samples of studies. The null hypotheses for the moderator analyses essentially states that the between-studies variance component (tau-squared) is zero.

**Geography**

Six geographical regions are considered as categories for the geography moderator based on the continents of the globe with the exception of grouping potential studies from the North American nations of Central America and Mexico with South American nations in the Latin American cluster. Studies conducted in the U.S. and Canada comprise the first category, whereas Latin American nations, European nations, Asian nations, African nations, and Australasian nations comprise the other five respective categories.
Implementation Scale

As a potential moderator of effect, implementation scale is treated as a categorical variable with the three levels of saturation. The scale of implementation refers to the degree to which an educational environment is saturated with Second Step implementation. For example, an entire school district comprised of multiple schools in which Second Step is implemented in every classroom is considered a highly saturated environment, regardless of the size of the district. In contrast, if only a small minority of classrooms within an entire district have implemented Second Step, it is characterized as low saturation, even if the raw number of students exposed to the program is greater than the number of students exposed in the high saturation context. It is the proportion of students exposed to the program versus those not exposed to the program that is instrumental in examining this aspect of the social ecological model of violence prevention. Simply using the number of students to indicate scale is not desired due to the variability in the number of students that can comprise a classroom, the number of classrooms that can be in a school, and the number of schools that can be in a district. For the sake of examining implementation scale within a social ecological model, the degree to which a student’s social ecology is saturated with Second Step lessons is more meaningful than the mere number of other students simultaneously exposed. In the present study, three levels of program implementation scale are constructed and defined as follows:

**High saturation:** Describes the scale in which every school in a district or multiple districts implementing Second Step in every classroom for which a grade-level curriculum exists. It also describes cases of studies in which the degree of district-wide
implementation was not reported and thus unknown, but multiple schools within a district implemented the program in every classroom.

Medium Saturation: Describes the scale in which at least half the schools in the district(s) implemented the program school-wide in the reported and known absence of district-wide implementation. It also describes cases in which neither the degree of district-wide implementation is reported or known, nor the degree of school-wide implementation is reported or known, but it is reported and known that multiple schools within a district implemented the program in multiple classrooms.

Low Saturation: Describes the scale in which it is reported and known that fewer than half the schools in a district or districts implemented the program.

Grade Level

Participant grade level comprises the final variable to be tested as a categorical moderator of effect. Grade level is treated as a categorical variable because the Second Step curriculum is separated and sold as grade level packages for preschool students, kindergarten through fifth grade students, and middle school students.

Training

Four levels of implementer training exist, including on-site training by authorized parties, access to online training materials provided by the publisher, a combination of these two, informal training provided by individuals not authorized by the publisher, and there is also the possibility that the program can be implemented without training.
Dependent Variable Reporter

Studies of Second Step may include data collected from different individuals, such as students providing self reports, teachers, parents, peers, and independent observers. The moderating impact of the reporter will also be tested.

Missing Data

Studies were only included in the meta-analysis if all necessary effect size data such as means, standard deviations, sample sizes, t-values, F-values, p-values, and effect size statistics were reported. No imputations of these forms of study data were conducted. In studies that included pre-test and post-test data in which correlations of pre-test and post-test scores were not reported, a correlation coefficient of 0.5 was imputed.

Calculating Summary Statistics within Studies

Some studies included in the meta-analysis reported multiple outcomes. In some cases, studies reported multiple outcomes within a single dependent variable (i.e., antisocial behavior, prosocial behavior, knowledge). In other cases, studies reported outcomes from multiple subgroups of participants. In order to prevent the overrepresentation of any particular sample, summary statistics were calculated for studies in which multiple outcomes were reported for the same sample within a dependent variable (Borenstein et al., 2009). Effect size data was entered into CMA to calculate the standardized differences in means and variances of each outcome measure within a study. The weighted means of standardized differences were calculated to represent a summary statistic for each study. These summary statistics were re-entered into a new CMA file in which one effect size was calculated for each study within each
of the three dependent variables. Studies which reported outcomes that fit into more than one of the dependent variable categories of the meta-analysis are represented only once within each dependent variable. Similar procedures were used to summarize effect size data within studies with respect to moderating variables such as grade level package and dependent variable reporter in which more than one source could exist in a single study.

**Publication Bias**

Published and unpublished studies of Second Step are included in the present meta-analysis. Using a funnel plot, each study is plotted with its standard error on the y-axis, and the standardized difference in means on the x-axis. The symmetry of this distribution about the overall mean effect size of all studies is used to detect possible publication bias. A trim and fill technique is used to determine whether the overall effect size would be altered if studies imputed to create a symmetrical distribution were included in the meta-analysis. In addition, the effect size of published studies is compared with the effect size of unpublished studies.

**Software Packages**

Mendeley Desktop was used for reference management and screening purposes. Coded study data including moderator variables was managed using MS Excel. Effect size data was coded directly into Comprehensive Meta-Analysis. All study data was imported into Comprehensive Meta-Analysis (Biostat) for statistical analyses and reporting.
CHAPTER IV

RESULTS

Study Inclusion

Results from online database searches were exported to Mendeley. Results from hand search were also exported to Mendeley. Table 3 shows the results of the search strategy.

Table 3. Sources of Retrieved Studies

<table>
<thead>
<tr>
<th>Bibliographic database/Source</th>
<th>Number of candidate studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProQuest Dissertations (English)</td>
<td>292</td>
</tr>
<tr>
<td>ProQuest Dissertations (Translations)</td>
<td>24</td>
</tr>
<tr>
<td>EBSCO (English)</td>
<td>324</td>
</tr>
<tr>
<td>EBSCO (Translations)</td>
<td>44</td>
</tr>
<tr>
<td>PsycArticles (English)</td>
<td>39</td>
</tr>
<tr>
<td>PsycArticles (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>PsycINFO (English)</td>
<td>201</td>
</tr>
<tr>
<td>PsycINFO (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>Hand search</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>932</strong></td>
</tr>
</tbody>
</table>

The 924 candidate studies from bibliographic databases were screened by Coder A and Coder B. Twenty percent of these (k=185) studies were screened for inclusion by both coders to ensure reliability of the screening procedures. Of these 185 studies, 125 were independently tagged for exclusion by both Coder A and Coder B, and 39 studies were independently tagged for inclusion by both Coder A and Coder B. Coder A identified one study for inclusion that Coder B excluded, and Coder B included 20
studies that Coder A excluded. Pre-consensus inter-rater reliability for screening was 165 of the 185 studies (89%).

Table 4. Study Screening Reliability

<table>
<thead>
<tr>
<th></th>
<th>Coder B inclusion</th>
<th>Coder B exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coder A inclusion</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>Coder A exclusion</td>
<td>20</td>
<td>125</td>
</tr>
</tbody>
</table>

Initially, Coder B held a broader interpretation of the inclusion criteria. This conservative approach led to the identification of more studies. At this stage, erring on the side of false positives is preferable to erring on the side of false negatives because full text screening would eliminate studies that did not meet inclusion criteria. All discrepancies were resolved after a consensus meeting to discuss reasons for inclusion and exclusion of studies, yielding a post-consensus screening reliability of 100% prior to screening of the remaining candidate studies. Results of the title and abstract screening are presented in Table 5.

Table 5. Title and Abstract Screening Results

<table>
<thead>
<tr>
<th>Bibliographic database/Source</th>
<th>Number of candidate studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProQuest Dissertations (English)</td>
<td>111</td>
</tr>
<tr>
<td>ProQuest Dissertations (Translations)</td>
<td>8</td>
</tr>
<tr>
<td>EBSCO (English)</td>
<td>38</td>
</tr>
<tr>
<td>EBSCO (Translations)</td>
<td>10</td>
</tr>
<tr>
<td>PsycArticles (English)</td>
<td>3</td>
</tr>
<tr>
<td>PsycArticles (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>PsycINFO (English)</td>
<td>51</td>
</tr>
<tr>
<td>PsycINFO (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>Hand search</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>229</td>
</tr>
</tbody>
</table>
Finally, a full text screening was conducted to identify a final set of studies for the meta-analysis. Reports were downloaded, and the coders searched the full texts to identify studies that met inclusion eligibility criteria. Results of the full text screening are shown in Table 6.

Table 6. Full Text Screening Results

<table>
<thead>
<tr>
<th>Bibliographic database/Source</th>
<th>Number of candidate studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProQuest Dissertations (English)</td>
<td>8</td>
</tr>
<tr>
<td>ProQuest Dissertations (Translations)</td>
<td>5</td>
</tr>
<tr>
<td>EBSCO (English)</td>
<td>8</td>
</tr>
<tr>
<td>EBSCO (Translations)</td>
<td>2</td>
</tr>
<tr>
<td>PsycArticles (English)</td>
<td>0</td>
</tr>
<tr>
<td>PsycArticles (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>PsycINFO (English)</td>
<td>13</td>
</tr>
<tr>
<td>PsycINFO (Translations)</td>
<td>0</td>
</tr>
<tr>
<td>Hand search</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
</tr>
</tbody>
</table>

Of these studies, six dissertations from both the hand search and database search were unavailable by author request, and two were excluded after contacting study authors who indicated that results from the same sample had been reported in other papers which met inclusion criteria. In two excluded studies, Second Step was included in a host of simultaneously implemented interventions, and it was not possible to separate the effects of this intervention alone. Four studies were excluded due to insufficient reporting of effect size data. Three other studies were excluded because only adjusted data was reported, and three were excluded because outcomes did not match the dependent variables of interest in the present study. An example of an excluded study at this final stage of study selection is the frequently cited study by Frey et al. (2005). It is a large
study on a sample of 1,253 participants. It was excluded on the grounds that the measured outcomes of interest (social competence and antisocial behavior) were reported as a function of students’ attribution biases and other covariates. Twenty-four studies are included in the present meta-analysis.

Figure 1. PRISMA Search Flow Diagram
Coding Study Data

Reliability

Coder A and Coder C coded eight studies to test for reliability. Out of 2,254 pieces of coded data within these eight studies, only 19 discrepancies existed, yielding an initial reliability coefficient of 0.99. The discrepancies arose out of a misinterpretation of two codes by Coder C. Upon discussion, this misinterpretation was corrected and post-consensus reliability was 100%. Coder A continued to code the remaining 16 studies.

Main Effects Analysis

Six separate meta-analyses were carried out for single-group design studies and two-group design studies with respect for each of the three dependent variables in order to determine overall program effects on antisocial behavior, prosocial behavior, and knowledge. In addition to these six main effects analyses, studies were combined across design types with respect to each of the three dependent variables in order to conduct moderator analyses. Results from the six main effects meta-analyses are described in the following sections. Table 7 shows the number of studies included in each of the six meta-analyses for main effects. Note the number of studies across columns and rows do not sum to the total number of studies in the meta-analysis because many studies include more than one outcome type.
Table 7. Studies and Participants by Design and Outcome

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Independent Group Design Studies</th>
<th>Single Group Design Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Studies</td>
<td>Number of Participants</td>
</tr>
<tr>
<td>Prosocial Outcomes</td>
<td>9</td>
<td>4,207</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td>10</td>
<td>6,427</td>
</tr>
<tr>
<td>Knowledge Outcomes</td>
<td>2</td>
<td>215</td>
</tr>
</tbody>
</table>

**Single Group Repeated Measures Designs**

Studies that were conducted as single group repeated measures designs are described in the following section. Studies that reported outcomes under more than one category are included in each of the relevant meta-analyses. Pre-post effect size correlations that were not reported in primary studies were imputed with a correlation coefficient of 0.5.

**Prosocial outcomes.** Based on the prosocial outcome measures reported in six studies, the standardized mean differences in pre-test to post-test scores ranged from -0.182 (Rosen, 2013) to 0.165 (McMahon, 2003). Using a random effects model, the overall standardized mean difference for all six studies is 0.090 (95% confidence interval: 0.044-0.136) with a standard error of 0.024. This represents a small, but significant overall effect with respect to the observed increase in prosocial outcomes associated with participation in the Second Step program. The Q-statistic associated with this effect size is 7.754 (df=5; p=0.170), $\tau^2$ is 0.001, and the null hypothesis of homogeneity of effect sizes cannot be rejected. This is the test of whether the between-studies variance ($\tau$-squared) significantly differs from zero. However, this does not necessarily mean the
effects are consistent. The power of this test may be low due to small and imprecise studies and the number of studies included. It is difficult to compute tau-squared with precision using a small number of studies (Borenstein et al., 2009). An $I^2$ value of 35.523% reflects the proportion of observed variance reflected by real differences in effect size. Figure 2 is a Forest plot summarizing this data.

### Main Effects: Prosocial Outcomes (Single Group Repeated Measures Designs)

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Outcome</th>
<th>Std diff in means</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooke</td>
<td>PROSOCIAL</td>
<td>0.074</td>
<td>0.016</td>
<td>0.000</td>
<td>0.042</td>
<td>0.106</td>
<td>4.571</td>
<td>0.000</td>
</tr>
<tr>
<td>Edwards</td>
<td>PROSOCIAL</td>
<td>0.103</td>
<td>0.032</td>
<td>0.001</td>
<td>0.039</td>
<td>0.166</td>
<td>3.163</td>
<td>0.002</td>
</tr>
<tr>
<td>McMahon, 2000</td>
<td>PROSOCIAL</td>
<td>0.188</td>
<td>0.035</td>
<td>0.002</td>
<td>0.067</td>
<td>0.262</td>
<td>3.034</td>
<td>0.001</td>
</tr>
<tr>
<td>McMahon, 2003</td>
<td>PROSOCIAL</td>
<td>0.185</td>
<td>0.030</td>
<td>0.001</td>
<td>0.050</td>
<td>0.262</td>
<td>3.304</td>
<td>0.001</td>
</tr>
<tr>
<td>Rosen</td>
<td>PROSOCIAL</td>
<td>0.162</td>
<td>0.017</td>
<td>0.002</td>
<td>0.043</td>
<td>0.101</td>
<td>-1.408</td>
<td>0.160</td>
</tr>
<tr>
<td>Tynes-Jones</td>
<td>PROSOCIAL</td>
<td>0.102</td>
<td>0.034</td>
<td>0.001</td>
<td>0.023</td>
<td>0.226</td>
<td>1.602</td>
<td>0.109</td>
</tr>
</tbody>
</table>

-0.50 -0.25 0.00 0.25 0.50

Less Prosocial More Prosocial

Random Effects

Figure 2. Forest Plot of Prosocial Outcomes for Single Group Repeated Measures Designs

**Antisocial outcomes.** Based on the antisocial outcome measures reported in six studies, the standardized mean differences in pre-test to post-test scores ranged from -0.113 (Cooke, 2007) to 0.163 (Tynes-Jones, 2006). Using a random effects model, the overall standardized mean difference for all six studies is 0.006 (95% confidence interval: -0.099-0.087) with a standard error of 0.047. This represents a non-significant overall effect with respect to the observed increase in antisocial outcomes associated with participation in the Second Step program. Note increased antisocial outcomes are coded as a negative effect. The Q-statistic associated with this effect size is 34.433 (df=5; p<0.001), tau$^2$ is 0.011, and the null hypothesis of homogeneity of effect sizes is rejected.
An $I^2$ value of 85.479% reflects the proportion of observed variance reflected by real differences in effect size. Figure 3 is a Forest plot summarizing this data.

**Main Effects: Antisocial Outcomes (Single Group Repeated Measures Designs)**

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome</th>
<th>Std</th>
<th>Standerr</th>
<th>Variance</th>
<th>Lower</th>
<th>Upper</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brow</td>
<td>ANTISOCIA -</td>
<td>0.07</td>
<td>0.00</td>
<td>0.04</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook</td>
<td>ANTISOCIA -</td>
<td>0.02</td>
<td>0.00</td>
<td>0.12</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edward</td>
<td>ANTISOCIA 0.04</td>
<td>0.04</td>
<td>0.00</td>
<td>1.67</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>House</td>
<td>ANTISOCIA -</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMahon</td>
<td>ANTISOCIA 0.10</td>
<td>0.04</td>
<td>0.00</td>
<td>0.18</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tynes</td>
<td>ANTISOCIA 0.16</td>
<td>0.09</td>
<td>0.00</td>
<td>0.34</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Random**

Figure 3. Forest Plot of Antisocial Outcomes for Single Group Repeated Measures Designs

**Knowledge outcomes.** Based on the knowledge outcome measures reported in eight studies, the standardized mean differences in pre-test to post-test scores ranged from -0.626 (Rosen) to 1.981 (Brown). Using a random effects model, the overall standardized mean difference for all six studies is 0.463 (95% confidence interval: 0.249-0.677) with a standard error of 0.109. This represents a moderately large significant overall effect with respect to the observed increase in knowledge outcomes associated with participation in the Second Step program. The Q-statistic associated with this effect size is 460.746 (df=7; p<0.001), $\tau^2$ is 0.089, and the null hypothesis of homogeneity of effect sizes is rejected. An $I^2$ value of 98.481% reflects the proportion of observed variance reflected by real differences in effect size. Figure 4 is a Forest plot summarizing this data.
Main Effects: Knowledge Outcomes (Single Group Repeated Measures Designs)

<table>
<thead>
<tr>
<th>Study name</th>
<th>Outcome</th>
<th>Std diff in means</th>
<th>Standard error</th>
<th>Variance Lower limit</th>
<th>Upper limit</th>
<th>z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>KNOWLEDGE</td>
<td>1.981</td>
<td>0.134</td>
<td>0.018</td>
<td>1.717</td>
<td>2.244</td>
<td>0.000</td>
</tr>
<tr>
<td>Edwards</td>
<td>KNOWLEDGE</td>
<td>0.286</td>
<td>0.047</td>
<td>0.052</td>
<td>0.175</td>
<td>0.377</td>
<td>5.708</td>
</tr>
<tr>
<td>Gruber</td>
<td>KNOWLEDGE</td>
<td>0.389</td>
<td>0.078</td>
<td>0.026</td>
<td>0.235</td>
<td>0.543</td>
<td>4.962</td>
</tr>
<tr>
<td>McMinn, 2002</td>
<td>KNOWLEDGE</td>
<td>0.119</td>
<td>0.002</td>
<td>0.000</td>
<td>0.115</td>
<td>0.124</td>
<td>54.463</td>
</tr>
<tr>
<td>McMinn, 2003</td>
<td>KNOWLEDGE</td>
<td>0.335</td>
<td>0.067</td>
<td>0.008</td>
<td>0.194</td>
<td>0.507</td>
<td>3.834</td>
</tr>
<tr>
<td>Neace</td>
<td>KNOWLEDGE</td>
<td>1.113</td>
<td>0.069</td>
<td>0.005</td>
<td>0.977</td>
<td>1.250</td>
<td>16.029</td>
</tr>
<tr>
<td>Rosen</td>
<td>KNOWLEDGE</td>
<td>-0.508</td>
<td>0.140</td>
<td>0.020</td>
<td>0.050</td>
<td>0.357</td>
<td>5.708</td>
</tr>
<tr>
<td>Sprague</td>
<td>KNOWLEDGE</td>
<td>0.199</td>
<td>0.026</td>
<td>0.001</td>
<td>0.148</td>
<td>0.249</td>
<td>7.702</td>
</tr>
</tbody>
</table>

Random Effects

Figure 4. Forest Plot of Knowledge Outcomes for Single Group Repeated Measures Designs

Independent Group Designs

Studies that were designed with separate treatment and comparison groups are described in the following section. Studies that reported outcomes under more than one category are included in each of the relevant meta-analyses. For example, notice that studies by Grumm (2012), Lipschutz (2010), Ruby (2010), and Taub (2002) are included in both analyses of prosocial outcomes and antisocial outcomes.

Prosocial outcomes. Based on the prosocial outcome measures reported in nine studies, the standardized mean differences between treatment and control group scores ranged from -0.239 (Grumm, 2013) to 0.502 (Beisly, 2011). Using a random effects model, the overall standardized mean difference for all nine studies is 0.061 (95% confidence interval: -0.020 - 0.142) with a standard error of 0.041. This represents a non-significant overall effect with respect to the observed increase in prosocial outcomes associated with participation in the Second Step program. The Q-statistic associated with this effect size is 21.390 (df=8; p=0.006), tau² is 0.008, and the null hypothesis of homogeneity of effect sizes is rejected. An I² value of 62.600% reflects the proportion of
observed variance reflected by real differences in effect size. Figure 5 is a Forest plot summarizing this data.

### Main Effects: Prosocial Outcomes (Group Designs)

<table>
<thead>
<tr>
<th>Study name</th>
<th>Outcome</th>
<th>Std diff in means</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beisly</td>
<td>PROSOCIAL</td>
<td>0.502</td>
<td>0.211</td>
<td>0.046</td>
<td>0.099</td>
<td>2.361</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>Grossman</td>
<td>PROSOCIAL</td>
<td>0.015</td>
<td>0.036</td>
<td>0.021</td>
<td>0.054</td>
<td>0.065</td>
<td>0.432</td>
<td></td>
</tr>
<tr>
<td>Grumm</td>
<td>PROSOCIAL</td>
<td>-0.239</td>
<td>0.105</td>
<td>0.011</td>
<td>-0.445</td>
<td>-0.024</td>
<td>2.288</td>
<td></td>
</tr>
<tr>
<td>Houtsen</td>
<td>PROSOCIAL</td>
<td>0.006</td>
<td>0.052</td>
<td>0.033</td>
<td>-0.017</td>
<td>0.188</td>
<td>1.640</td>
<td></td>
</tr>
<tr>
<td>Lippschutz</td>
<td>PROSOCIAL</td>
<td>0.222</td>
<td>0.259</td>
<td>0.067</td>
<td>-0.285</td>
<td>0.730</td>
<td>0.859</td>
<td></td>
</tr>
<tr>
<td>Oudemond</td>
<td>PROSOCIAL</td>
<td>0.006</td>
<td>0.046</td>
<td>0.022</td>
<td>-0.863</td>
<td>0.095</td>
<td>1.314</td>
<td></td>
</tr>
<tr>
<td>Ruby</td>
<td>PROSOCIAL</td>
<td>0.098</td>
<td>0.059</td>
<td>0.033</td>
<td>-0.017</td>
<td>0.213</td>
<td>1.666</td>
<td></td>
</tr>
<tr>
<td>Schick</td>
<td>PROSOCIAL</td>
<td>0.007</td>
<td>0.070</td>
<td>0.050</td>
<td>-0.040</td>
<td>0.233</td>
<td>1.389</td>
<td></td>
</tr>
<tr>
<td>Taub</td>
<td>PROSOCIAL</td>
<td>0.066</td>
<td>0.253</td>
<td>0.064</td>
<td>0.140</td>
<td>1.120</td>
<td>2.514</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.061</td>
<td>0.041</td>
<td>0.032</td>
<td>-0.020</td>
<td>0.140</td>
<td>1.486</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Forest Plot of Prosocial Outcomes for Independent Group Designs

**Antisocial outcomes.** Based on the antisocial outcome measures reported in ten studies, the standardized mean differences between treatment and control group scores ranged from -0.512 (Grumm, 2013) to 0.519 (Taub, 2002). Using a random effects model, the overall standardized mean difference for all six studies is -0.113 (95% confidence interval: -0.238- 0.012) with a standard error of 0.064. This represents a non-significant overall effect with respect to the observed increase in antisocial outcomes associated with participation in the Second Step program. Note increased antisocial outcomes are coded as a negative effect. The Q-statistic associated with this effect size is 129.173 (df=9; p<0.001), tau² is 0.026, and the null hypothesis of homogeneity of effect sizes is rejected. An I² value of 93.032% reflects the proportion of observed variance reflected by real differences in effect size. Figure 6 is a Forest plot summarizing this data.
Figure 6. Forest Plot of Antisocial Outcomes for Independent Group Designs

Knowledge outcomes. Based on the knowledge outcome measures reported in two studies, the standardized mean differences between treatment and control group scores ranged from 0.490 (Lipschutz, 2010) to 0.968 (Hart, 2009). Using a random effects model, the overall standardized mean difference for both studies is 0.767 (95% confidence interval: 0.304-1.230) with a standard error of 0.236. This represents a large significant overall effect with respect to the observed increase in knowledge outcomes associated with participation in the Second Step program. The Q-statistic associated with this effect size is 2.263 (df=1; p=0.133), tau² is 0.064, and the null hypothesis of homogeneity of effect sizes cannot be rejected. However, this does not necessarily mean the effects are consistent. The effect sizes of these two studies are nearly 0.5 standard deviations apart. The power of this test may be low due the inclusion of only two studies. It is difficult to compute tau-squared with precision using a small number of studies (Borenstein et al., 2009). An I² value of 55.810% reflects the proportion of observed variance reflected by real differences in effect size. Figure 7 is a Forest plot summarizing this data.
Figure 7. Forest Plot of Knowledge Outcomes for Independent Group Designs

Table 8. Main Effect Homogeneity Tests

<table>
<thead>
<tr>
<th>Outcome</th>
<th>ES</th>
<th>p-value</th>
<th>SE</th>
<th>tau-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Group</td>
<td>0.09</td>
<td>&lt;0.001</td>
<td>0.002</td>
<td>0.032</td>
</tr>
<tr>
<td>Independent Groups</td>
<td>0.061</td>
<td>0.137</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Group</td>
<td>-0.006</td>
<td>0.904</td>
<td>0.009</td>
<td>0.011</td>
</tr>
<tr>
<td>Independent Groups</td>
<td>-0.113</td>
<td>0.077</td>
<td>0.024</td>
<td>0.026</td>
</tr>
<tr>
<td>Knowledge Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Group</td>
<td>0.463</td>
<td>&lt;0.001</td>
<td>0.093</td>
<td>0.089</td>
</tr>
<tr>
<td>Independent Groups</td>
<td>0.767</td>
<td>0.001</td>
<td>0.162</td>
<td>0.064</td>
</tr>
</tbody>
</table>

The $I^2$ values associated with each of these six meta-analyses indicate the percentages of the observed variability among studies due to factors other than sampling error. The $I^2$ values ranged from 35.523% to 98.481%. The following moderator analyses serve as an attempt to systematically model some of this variability based on variables derived from the social ecological model of youth violence prevention. Single group and independent group findings were relatively consistent with regards to magnitude and direction within each dependent variable with respect to study design. In order to increase the statistical power related to small study numbers within each
dependent variable, studies that were previously separated by study design were combined for moderator analyses. That is, both single group and independent group studies were analyzed together for each of the moderators analyses under prosocial, antisocial, and knowledge outcomes, respectively.

**Moderator Analyses**

Categorical moderator analyses were conducted using a random effects model and pooled estimates of variance across subgroups. There is reason to believe that variances at different levels of the moderator categories differ from one another. However, basing estimates of the separate variances on small samples of studies within each level is imprudent. A more conservative approach is to use a pooled variance estimate for the practical reason generated by small cell sizes within levels of the moderators. The rationale for applying a random effects model in all the moderator analyses is the same as applying it to the main effects meta-analyses. There is no reason to believe that all the studies are identical, even though the moderator analysis is an attempt to model some of the between-studies variance. The results of these moderator analyses are reported in the following section with respect to each of the three dependent variables and irrespective of study design. In other words, single group and independent group design studies were combined within each of the dependent variable categories, and moderator analyses were conducted on studies combined across designs. For each moderator variable (e.g., implementation scale), a statistically significant Q-statistic serves as evidence to reject the null hypothesis that there is no difference between the different categories or levels of the moderator (e.g., high-scale, medium-scale, low scale). A rejection of the null
hypothesis would indicate significant moderator effects by the variable under examination.

**Geography**

Despite a search that would include studies of Second Step and its international translations all over the world, identified studies were returned from only two regions – Europe and USA/Canada. Therefore, these two categories comprised the extent of the moderator analysis instead of the originally intended six. Of the European studies, two were conducted in Germany, and one was conducted in Norway. Analysis of geography as a moderator was limited only to prosocial and antisocial outcomes. All studies reporting knowledge outcomes were conducted in USA/Canada. Table 9 shows the number of studies included in the analyses of geography as a moderator.

Table 9. Studies Included in Geography Moderator Analysis

<table>
<thead>
<tr>
<th>Region</th>
<th>Prosocial Outcomes</th>
<th>Antisocial Outcomes</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA/Canada</td>
<td>12*</td>
<td>14*</td>
<td>10</td>
</tr>
<tr>
<td>Europe</td>
<td>3*</td>
<td>2*</td>
<td>0</td>
</tr>
<tr>
<td>Latin America</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Africa</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Australasia</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*included in moderator analysis

On prosocial outcomes, 12 studies conducted in the USA/Canada produced an effect size of 0.081 (95% confidence interval: 0.023-0.116) with a standard error of 0.026. This represents a small significant (p=0.002) effect with respect to the observed increase in prosocial outcomes. The three European studies produced an effect of 0.025 (95% confidence interval: -0.080-0.130) with a standard error of 0.054. This represents a non-significant effect (p=0.642). The between groups Q statistic is 0.864 (df=1;
p=0.353), and the null hypothesis that studies conducted in Europe and studies conducted in USA/Canada measuring prosocial outcomes are homogeneous cannot be rejected. However, a distinction can be made that studies conducted in USA/Canada produced a statistically significant positive effect in increasing prosocial outcomes, whereas confidence interval of effect sizes associated with European studies includes zero and is statistically non-significant. Figure 8 is a Forest plot summarizing this data.

![Forest Plot of Geography Moderator on Prosocial Outcomes](image)

**Figure 8. Forest Plot of Geography Moderator on Prosocial Outcomes**

On antisocial outcomes, 14 studies conducted in the USA/Canada produced an effect size of -0.034 (95% confidence interval: -0.112- 0.043) with a standard error of 0.039. This represents a non-significant (p=0.382) effect with respect to the observed increase in antisocial outcomes. The two European studies produced an effect of -0.363 (95% confidence interval: -0.596- -0.130) with a standard error of 0.119. This represents a significant effect (p=0.002) with respect to the observed increase in antisocial outcomes. The between-groups Q statistic is 6.878 (df=1; p=0.009), and the null hypothesis that studies conducted in Europe and studies conducted in USA/Canada...
measuring antisocial outcomes is rejected. Underscoring the statistically significant
test statistic, the contrast between the non-significant effect size of studies
carried out in USA/Canada and the significant effect size of European studies measuring
antisocial outcomes should be noted. Figure 9 is a Forest plot summarizing this data.

![Moderator: Geography (Antisocial)](image)

Figure 9. Forest Plot of Geography Moderator on Antisocial Outcomes

Based on the between-groups Q-statistic, which is an indicator of the differences
between the effect sizes moderated by geography, it does not appear that geography has a
significant moderating effect on Second Step’s impact on prosocial associated
outcomes. However, there does appear to be a meaningful difference between Second
Step’s effect in USA/Canada and its effect in Europe with regards to antisocial outcomes.
European studies of Second Step’s effectiveness indicate no significant effect on
prosocial outcomes and a significant increase in antisocial outcomes after program
implementation, whereas studies conducted in USA/Canada showed a statistically
significant increase in prosocial outcomes and a non-significant effect on students’
antisocial outcomes. Overall, it appears that Second Step has a relatively beneficial
effect on students in USA/Canada compared to those in Europe although this difference may not bear statistical significance.

**Implementation Scale**

Three levels of implementation scale were examined on prosocial and antisocial outcomes. On knowledge outcomes, only two levels of implementation scale were examined. The ‘medium’ implementation scale was only represented by one study (Brown) for knowledge outcomes and removed. Table 10 shows the number of studies and participants by implementation scale and outcome.

**Table 10. Studies Included in Implementation Scale Moderator Analysis**

<table>
<thead>
<tr>
<th>Implementation Scale</th>
<th>Prosocial Outcomes</th>
<th>Antisocial Outcomes</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of Studies</td>
<td>Number of Participants</td>
<td>Number of Studies</td>
</tr>
<tr>
<td>High Saturation</td>
<td>3*</td>
<td>2,252</td>
<td>5*</td>
</tr>
<tr>
<td>Medium Saturation</td>
<td>4*</td>
<td>1,070</td>
<td>4*</td>
</tr>
<tr>
<td>Low Saturation</td>
<td>8*</td>
<td>3,279</td>
<td>7*</td>
</tr>
</tbody>
</table>

*included in moderator analyses

On prosocial outcomes, three studies in which Second Step was implemented at a high scale produced an effect size of 0.065 (95% confidence interval: -0.018 - 0.147) with a standard error of 0.042. This represents a non-significant (p=0.124) effect with respect to the observed increase in prosocial outcomes. The four studies in which Second Step was implemented at a medium scale produced an effect of 0.156 (95% confidence interval: 0.038 - 0.275) with a standard error of 0.060. This represents a small, but significant effect (p=0.010) with respect to the observed increase in prosocial outcomes. The eight studies in which Second Step was implemented at a low scale produced an
effect of 0.041 (95% confidence interval: -0.031 - 0.113) with a standard error of 0.037. This represents a non-significant effect (p=0.259) with respect to the observed increase in prosocial outcomes. The between-groups Q statistic is 2.668(df=2; p=0.263), and the null hypothesis that studies there is no difference between levels of implementation scale cannot be rejected for prosocial outcomes. However, medium-scale implementation produced a significant positive effect on increasing prosocial outcomes, whereas high-scale and low-scale implementation produced a non-significant effect on prosocial outcomes. Figure 10 is a Forest plot summarizing this data.

Figure 10. Forest Plot of Implementation Scale Moderator on Prosocial Outcomes

On antisocial outcomes, five studies in which Second Step was implemented at a high scale produced an effect size of -0.119 (95% confidence interval: -0.241 - 0.002) with a standard error of 0.062. This represents a non-significant (p=0.054) effect with respect to the observed increase in antisocial outcomes. The four studies in which Second Step was implemented at a medium scale produced an effect of -0.032 (95% confidence interval: -0.202 - 0.137) with a standard error of 0.087. This represents a non-
significant effect ($p=0.708$) with respect to the observed increase in antisocial outcomes. The seven studies in which Second Step was implemented at a low scale produced an effect of -0.022 (95% confidence interval: -0.155 - 0.111) with a standard error of 0.068. This represents a non-significant effect ($p=0.749$) with respect to the observed increase in antisocial outcomes. The between-groups Q statistic is 1.314 (df=2; $p=0.518$), and the null hypothesis that there is no difference between levels of implementation scale cannot be rejected for antisocial outcomes. Neither high-scale, medium-scale, nor low-scale implementation produced statistically significant effects on antisocial outcomes. Figure 11 is a Forest plot summarizing this data.

<table>
<thead>
<tr>
<th>Study name</th>
<th>Outcome</th>
<th>Statistics for each study</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooke</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.113</td>
</tr>
<tr>
<td>Edwards</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.040</td>
</tr>
<tr>
<td>Espelage</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.070</td>
</tr>
<tr>
<td>Hanks</td>
<td>ANTI</td>
<td>low</td>
<td>0.080</td>
</tr>
<tr>
<td>Carlton</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.040</td>
</tr>
<tr>
<td>Seligman</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.040</td>
</tr>
<tr>
<td>Grossman</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.050</td>
</tr>
<tr>
<td>Gurria</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.050</td>
</tr>
<tr>
<td>Luchak</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.050</td>
</tr>
<tr>
<td>McGee</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.050</td>
</tr>
<tr>
<td>Mckean</td>
<td>ANTI</td>
<td>low</td>
<td>0.150</td>
</tr>
<tr>
<td>Tynes-Jones</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.150</td>
</tr>
<tr>
<td>Tomlin</td>
<td>ANTI</td>
<td>low</td>
<td>0.050</td>
</tr>
<tr>
<td>Tynes-Jones</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.150</td>
</tr>
<tr>
<td>Overall</td>
<td>ANTISOCIAL</td>
<td>low</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Figure 11. Forest Plot of Implementation Scale Moderator on Antisocial Outcomes

On knowledge outcomes, high scale of implementation was compared to low scale of implementation only. The four studies in which Second Step was implemented at a high scale produced an effect size of .488 (95% confidence interval: 0.143-0.834) with a standard error of 0.176. This represents a significant ($p=0.006$) effect with respect
to the observed increase in student knowledge. The five studies in which Second Step was implemented at a low scale produced an effect of 0.230 (95% confidence interval: -0.102 - 0.562) with a standard error of 0.169. This represents a non-significant effect (p=0.174) with respect to the observed increase in student knowledge. The between-groups Q statistic is 1.117 (df=1; p=0.291), and the null hypothesis that studies there is no difference between levels of implementation scale cannot be rejected for student knowledge. Figure 12 is a Forest plot summarizing this data.

<table>
<thead>
<tr>
<th>Study name</th>
<th>Outcome</th>
<th>Statistics for each study</th>
<th>Std diff in means</th>
<th>95% confidence limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edwards</td>
<td>KNOWLEDGE</td>
<td>0.366</td>
<td>0.467</td>
<td>0.056</td>
<td>0.175</td>
<td>0.537</td>
</tr>
<tr>
<td>Gruber</td>
<td>KNOWLEDGE</td>
<td>0.389</td>
<td>0.379</td>
<td>0.086</td>
<td>0.325</td>
<td>0.543</td>
</tr>
<tr>
<td>Herre</td>
<td>KNOWLEDGE</td>
<td>1.113</td>
<td>0.959</td>
<td>0.185</td>
<td>0.977</td>
<td>1.330</td>
</tr>
<tr>
<td>Sorensen</td>
<td>KNOWLEDGE</td>
<td>0.180</td>
<td>0.066</td>
<td>0.091</td>
<td>0.194</td>
<td>0.269</td>
</tr>
<tr>
<td>Hart</td>
<td>KNOWLEDGE</td>
<td>0.968</td>
<td>0.170</td>
<td>0.031</td>
<td>0.143</td>
<td>0.354</td>
</tr>
<tr>
<td>Lomax</td>
<td>KNOWLEDGE</td>
<td>0.480</td>
<td>0.262</td>
<td>0.110</td>
<td>0.353</td>
<td>0.607</td>
</tr>
<tr>
<td>Malott, 2009</td>
<td>KNOWLEDGE</td>
<td>0.135</td>
<td>0.057</td>
<td>0.080</td>
<td>0.104</td>
<td>0.167</td>
</tr>
<tr>
<td>Rosen</td>
<td>KNOWLEDGE</td>
<td>0.280</td>
<td>0.188</td>
<td>0.180</td>
<td>0.103</td>
<td>0.352</td>
</tr>
</tbody>
</table>

Overall effect: 0.354 ± 0.129 (Z = 2.745, p = 0.006)

Figure 12. Forest Plot of Implementation Scale Moderator on Knowledge Outcomes

Based on these data between groups Q-statistic, which is an indicator of the differences between the effect sizes moderated by implementation scale, it does not appear that implementation scale has a significant moderating effect on Second Step’s impact on prosocial outcomes, antisocial outcomes, or student knowledge related to violence prevention. Although the p-value of the between-groups Q-statistic was not statistically significant at the 0.05 probability level, further examination of the effect size data does indicate that medium-scale implementation produced a statistically significant

Moderator: Implementation Scale (Knowledge)
positive effect on prosocial outcomes and high-scale implementation produced a statistically significant positive effect on knowledge outcomes.

**Training**

Several different modes of training were reported in the primary studies. These modes were collapsed into two categories: training obtained through authorized parties and materials provided by Second Step’s publisher; and all other forms of training, including studies that did not specifically report their training procedures. Table 11 shows the number of studies by training mode and outcome.

Table 11. Studies Included in Training Moderator Analysis

<table>
<thead>
<tr>
<th>Training Mode</th>
<th>Prosocial Outcomes</th>
<th>Antisocial Outcomes</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certified Onsite</td>
<td>8*</td>
<td>7*</td>
<td>3*</td>
</tr>
<tr>
<td>Certified Online</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Certified Combined</td>
<td>0</td>
<td>0</td>
<td>1*</td>
</tr>
<tr>
<td>Informal Reported</td>
<td>2*</td>
<td>3*</td>
<td>3*</td>
</tr>
<tr>
<td>Not Reported</td>
<td>5*</td>
<td>5*</td>
<td>3*</td>
</tr>
</tbody>
</table>

*included in moderator analysis

On prosocial outcomes, eight studies in which implementers received training through modes approved and authorized by Second Step’s publisher produced an effect size of 0.050 (95% confidence interval: -0.011- 0.111) with a standard error of 0.031. This represents a non-significant (p=0.107) effect with respect to the observed increase in prosocial outcomes. The seven studies in which approved or authorized trainings were not reported produced an effect of 0.097 (95% confidence interval: 0.025- 0.169) with a standard error of 0.037. This represents a small, but significant effect (p=0.009). The between groups Q statistic is 0.939 (df=1; p=0.333), and the null hypothesis that studies measuring prosocial outcomes conducted with approved and authorized training and
those that did not report approved or authorized trainings are homogeneous cannot be rejected. However, studies in which approved or authorized trainings were not reported produced a significant positive effect, whereas those reporting approved or authorized trainings did not. Figure 13 is a Forest plot summarizing this data.

![Forest Plot of Training Moderator on Prosocial Outcomes](image)

**Figure 13. Forest Plot of Training Moderator on Prosocial Outcomes**

On antisocial outcomes, eight studies in which implementers received training through modes approved and authorized by Second Step’s publisher produced an effect size of -0.091 (95% confidence interval: -0.190- 0.008) with a standard error of 0.051. This represents a non-significant (p=0.071) effect with respect to the observed increase in antisocial outcomes. The eight studies in which approved or authorized trainings were not reported produced an effect of -0.035 (95% confidence interval: -0.148- 0.077) with a standard error of 0.057. This represents a non-significant effect (p=0.541) with respect to the observed increase in antisocial outcomes. The between groups Q statistic is 0.536 (df=1; p=0.464), and the null hypothesis that studies measuring antisocial outcomes conducted with approved and authorized training and those that did not report approved
or authorized trainings are homogeneous cannot be rejected. Neither group of studies based on training modality produced significant effect sizes on antisocial outcomes. Figure 14 is a Forest plot summarizing this data.

**Moderator: Training (Antisocial)**

<table>
<thead>
<tr>
<th>Study name</th>
<th>Outcome</th>
<th>Std diff in means</th>
<th>Std error</th>
<th>Standard error</th>
<th>Variance</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooke</td>
<td>ANTISOCIAL</td>
<td>0.013</td>
<td>0.009</td>
<td>0.029</td>
<td>0.152</td>
<td>-0.079</td>
<td>0.009</td>
<td>0.519</td>
<td>0.519</td>
</tr>
<tr>
<td>Espelage</td>
<td>ANTISOCIAL</td>
<td>0.053</td>
<td>0.003</td>
<td>0.053</td>
<td>0.079</td>
<td>-0.020</td>
<td>0.020</td>
<td>1.731</td>
<td>0.083</td>
</tr>
<tr>
<td>Grossen</td>
<td>ANTISOCIAL</td>
<td>0.025</td>
<td>0.001</td>
<td>0.025</td>
<td>0.078</td>
<td>0.021</td>
<td>1.031</td>
<td>0.308</td>
<td>0.001</td>
</tr>
<tr>
<td>Hunsley</td>
<td>ANTISOCIAL</td>
<td>0.074</td>
<td>0.002</td>
<td>0.047</td>
<td>0.171</td>
<td>0.014</td>
<td>1.673</td>
<td>1.059</td>
<td>0.001</td>
</tr>
<tr>
<td>McCabe</td>
<td>ANTISOCIAL</td>
<td>0.120</td>
<td>0.044</td>
<td>0.022</td>
<td>0.039</td>
<td>0.011</td>
<td>0.544</td>
<td>1.914</td>
<td>0.056</td>
</tr>
<tr>
<td>Shepard</td>
<td>ANTISOCIAL</td>
<td>0.228</td>
<td>0.001</td>
<td>0.032</td>
<td>0.457</td>
<td>-0.301</td>
<td>-2.023</td>
<td>0.309</td>
<td>0.000</td>
</tr>
<tr>
<td>Taylor</td>
<td>ANTISOCIAL</td>
<td>0.090</td>
<td>0.002</td>
<td>0.048</td>
<td>0.084</td>
<td>0.006</td>
<td>1.000</td>
<td>0.312</td>
<td>0.001</td>
</tr>
<tr>
<td>Turk</td>
<td>ANTISOCIAL</td>
<td>0.519</td>
<td>0.063</td>
<td>0.251</td>
<td>0.027</td>
<td>1.012</td>
<td>2.068</td>
<td>0.090</td>
<td>0.009</td>
</tr>
<tr>
<td>Brown</td>
<td>ANTISOCIAL</td>
<td>0.091</td>
<td>0.003</td>
<td>0.051</td>
<td>0.190</td>
<td>0.009</td>
<td>1.822</td>
<td>0.971</td>
<td>0.000</td>
</tr>
<tr>
<td>Edwards</td>
<td>ANTISOCIAL</td>
<td>0.212</td>
<td>0.046</td>
<td>0.214</td>
<td>0.308</td>
<td>0.632</td>
<td>0.988</td>
<td>0.323</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.112</td>
<td>0.006</td>
<td>0.078</td>
<td>0.263</td>
<td>0.044</td>
<td>1.399</td>
<td>0.163</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.043</td>
<td>0.002</td>
<td>0.040</td>
<td>0.035</td>
<td>0.100</td>
<td>0.972</td>
<td>0.264</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.512</td>
<td>0.012</td>
<td>0.111</td>
<td>0.729</td>
<td>-0.294</td>
<td>-4.658</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.053</td>
<td>0.007</td>
<td>0.289</td>
<td>0.441</td>
<td>0.572</td>
<td>1.052</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.044</td>
<td>0.002</td>
<td>0.043</td>
<td>0.018</td>
<td>0.167</td>
<td>2.867</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.112</td>
<td>0.015</td>
<td>0.012</td>
<td>0.439</td>
<td>0.004</td>
<td>1.670</td>
<td>0.094</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.163</td>
<td>0.008</td>
<td>0.080</td>
<td>0.013</td>
<td>0.240</td>
<td>1.816</td>
<td>0.063</td>
<td>0.000</td>
</tr>
<tr>
<td>Headrick</td>
<td>ANTISOCIAL</td>
<td>0.035</td>
<td>0.003</td>
<td>0.057</td>
<td>0.148</td>
<td>0.077</td>
<td>0.612</td>
<td>0.541</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall</td>
<td>ANTISOCIAL</td>
<td>0.087</td>
<td>0.001</td>
<td>0.038</td>
<td>0.141</td>
<td>0.008</td>
<td>1.758</td>
<td>0.079</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Random Effects**

Figure 14. Forest Plot of Training Moderator on Antisocial Outcomes

On knowledge outcomes, four studies in which implementers received training through modes approved and authorized by Second Step’s publisher produced an effect size of 0.262 (95% confidence interval: -0.208- 0.731) with a standard error of 0.240. This represents a non-significant (p=0.275) effect with respect to the observed increase in knowledge outcomes. The six studies in which approved or authorized trainings were not reported produced an effect of 0.696 (95% confidence interval: 0.302- 1.089) with a standard error of 0.201. This represents a moderately large significant effect (p=0.002).

The between groups Q statistic is 1.930 (df=1; p=0.165), and the null hypothesis that studies measuring knowledge outcomes conducted with approved and authorized training and those that did not report approved or authorized trainings are homogeneous cannot be rejected. Figure 15 is a Forest plot summarizing this data.
Based on the non-significant between-groups Q-statistic, it does not appear that training mode has a significant moderating effect on Second Step’s impact on prosocial outcomes, antisocial outcomes, or student knowledge related to violence prevention. However, based on prosocial and knowledge outcomes, effect sizes were positive and statistically significant in studies in which training was not reported or in which training was not provided using authorized Second Step materials, whereas those describing the use of authorized materials and/or parties in implementer training did not produce significant positive effect sizes.

**Dependent Variable Reporter**

The majority of studies employed self-reported and teacher-reported outcomes. Too few studies included peer, parent, or independent observer data to meaningfully include these as moderators in this analysis. Therefore, these outcomes were removed prior to conducting the moderator analysis, which only included self-reported data and teacher-reported data. Some studies relied on both teacher report and student report. These data are included in each of the levels of the moderator. Knowledge outcomes
were collected only using self-report measures. Table 12 shows the number of studies by DV reporter and outcome.

Table 12. Studies Included in Dependent Variable Reporter Moderator Analysis

<table>
<thead>
<tr>
<th>DV Reporter</th>
<th>Prosocial Outcomes</th>
<th>Antisocial Outcomes</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>9*</td>
<td>7*</td>
<td>10</td>
</tr>
<tr>
<td>Teacher</td>
<td>6*</td>
<td>11*</td>
<td>0</td>
</tr>
<tr>
<td>Parent/Caregiver</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Peer</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Independent Observer</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

*included in moderator analysis

On prosocial outcomes, nine studies that relied on student self-reports produced an effect size of 0.057 (95% confidence interval: -0.023- 0.138) with a standard error of 0.041. This represents a non-significant (p=0.164) effect with respect to the self-reported increase in prosocial outcomes. The six studies that relied on teacher reports produced an effect of 0.088 (95% confidence interval: 0.028- 0.205) with a standard error of 0.060. This represents a non-significant effect (p=0.137) with respect to prosocial outcomes.

The between groups Q statistic is 0.185 (df=1; p=0.667), and the null hypothesis there is no difference between self-reported and teacher reported prosocial outcomes cannot be rejected. Neither self-reported data nor teacher-reported data indicated significant effects of Second Step on prosocial outcomes. Figure 16 is a Forest plot summarizing this data.
Figure 16. Forest Plot of Dependent Variable Reporter Moderator on Prosocial Outcomes

On antisocial outcomes, seven studies that relied on student self-reports produced an effect size of -0.117 (95% confidence interval: -0.223 - 0.000) with a standard error of 0.059. This represents a small, but significant ($p=0.049$) effect with respect to the self-reported increase in antisocial outcomes. The 11 studies that relied on teacher reports produced an effect of -0.066 (95% confidence interval: -0.176 - 0.043) with a standard error of 0.056. This represents a non-significant effect ($p=0.236$) with respect to antisocial outcomes. The between groups Q statistic is 0.382 (df=1; $p=0.537$), and the null hypothesis there is no difference between self-reported and teacher reported antisocial outcomes cannot be rejected. However, a significant negative effect on antisocial outcomes was indicated by student self-report, whereas teacher-reported data indicated no significant effect. Figure 17 is a Forest plot summarizing this data.
Based on the between-groups Q-statistic, which is an indicator of the differences between the effect sizes moderated by dependent variable reporter, it does not appear that the reporter has a significant moderating effect on Second Step’s impact on prosocial outcomes or antisocial outcomes. However, there does appear to be a meaningful difference between teachers’ perceptions of student antisociality and the perception of students, who viewed themselves as more antisocial after participation in the program.

**Grade Level Package**

An overwhelming majority of the studies included in the present meta-analysis evaluated the outcomes of the Elementary (K-5) grade level curricula. An insufficient number of studies evaluated preschool or middle school curricula and reported these data separately by grade level. Therefore, a moderator analysis of grade level package was not conducted. Table 13 shows the number of studies by grade level package and outcome.
Table 13. Studies by Grade Level Package and Outcome

<table>
<thead>
<tr>
<th></th>
<th>Prosocial Outcomes</th>
<th>Antisocial Outcomes</th>
<th>Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preschool</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Elementary (K-5)</td>
<td>9</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Middle (6-8)</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Based on the moderator analyses described above neither implementation scale, training mode, or dependent variable reporter had a significant moderating effect on effect size. However, geography moderated effect on antisocial outcomes. The following table summarizes the results of the moderator analyses.

Table 14. Moderator Homogeneity Tests

<table>
<thead>
<tr>
<th></th>
<th>tau-squared</th>
<th>SE</th>
<th>Q between</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosocial Outcomes</td>
<td>0.003</td>
<td>0.003</td>
<td>0.864</td>
<td>0.353</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td>0.016</td>
<td>0.011</td>
<td>6.878</td>
<td>0.009</td>
</tr>
<tr>
<td>Implementation Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosocial Outcomes</td>
<td>0.003</td>
<td>0.003</td>
<td>2.668</td>
<td>0.263</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td>0.016</td>
<td>0.011</td>
<td>1.314</td>
<td>0.518</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosocial Outcomes</td>
<td>0.003</td>
<td>0.003</td>
<td>0.939</td>
<td>0.333</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td>0.016</td>
<td>0.011</td>
<td>0.536</td>
<td>0.464</td>
</tr>
<tr>
<td>Knowledge Outcomes</td>
<td>0.091</td>
<td>0.094</td>
<td>1.93</td>
<td>0.165</td>
</tr>
<tr>
<td>Dependent Variable Reporter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosocial Outcomes</td>
<td>0.005</td>
<td>0.004</td>
<td>0.185</td>
<td>0.667</td>
</tr>
<tr>
<td>Antisocial Outcomes</td>
<td>0.018</td>
<td>0.013</td>
<td>0.382</td>
<td>0.537</td>
</tr>
</tbody>
</table>
Publication Bias

Prosocial Outcomes

The distribution of the observed effect size values appear biased to the right of the overall standardized difference in means (0.070). By using Duval and Tweedie’s trim and fill to include two hypothetical studies that would balance out the distribution of effect sizes on this outcome, the overall standardized mean effect would be adjusted by -0.011 (0.059). A comparison of the mean effect sizes of published journal articles (0.076) to dissertations (0.048) yields a between-groups Q-statistic of 0.230 (df=1; p=0.632) which is statistically non-significant. Figure 18 is a funnel plot representing studies by their effect size (x-axis) and standard error of the effect size (y-axis). Observed studies are represented by empty circles, and imputed studies are represented by filled (black) circles.

Figure 18. Funnel Plot of Prosocial Outcomes by Report Type
Table 15. Trim and Fill Adjustments on Studies Reporting Prosocial Outcomes

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Study Name</th>
<th>Outcome</th>
<th>Std diff in means and 95% CI</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissertations</td>
<td>Beisly</td>
<td>PROSOCIAL</td>
<td>0.502, 0.211</td>
<td>0.045, 0.916</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Lapachoski</td>
<td>PROSOCIAL</td>
<td>0.502, 0.211</td>
<td>0.045, 0.916</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Osmundson</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Rauvan</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Tyres-Jones</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Gay</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Cooke</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Edwards</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Grossman</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Holsen</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>McMahon, 2000</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>McMahon, 2003</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Ruby</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Schick</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Taub</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Dissertations</td>
<td>Overall</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Cooke</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Edwards</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Grossman</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Holsen</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>McMahon, 2000</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>McMahon, 2003</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Ruby</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Schick</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Taub</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
<tr>
<td>Journals</td>
<td>Overall</td>
<td>PROSOCIAL</td>
<td>0.006, 0.045</td>
<td>0.002, 0.069</td>
</tr>
</tbody>
</table>

Figure 19. Forest Plot of Prosocial Outcomes by Report Type

**Antisocial Outcomes**

The distribution of the observed effect size values appear biased to the right of the overall standardized difference in means (-0.067). By using Duval and Tweedie’s trim and fill to include two hypothetical studies that would balance out the distribution of effect sizes on this outcome, the overall standardized mean effect would be adjusted by -0.018 (-0.084). A comparison of the mean effect size of published journal articles (-0.055) to dissertations (-0.125) yields a between-groups Q-statistic of 0.768 (df=1; p=0.381) which is statistically non-significant.
Table 16. Trim and Fill Adjustments for Antisocial Outcomes

<table>
<thead>
<tr>
<th>Study Name</th>
<th>Outcome</th>
<th>Statistics for each study Std diff in means</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissertation Bogue</td>
<td>ANTISOCIAL</td>
<td>0.212</td>
<td>0.046</td>
</tr>
<tr>
<td>Dissertation Lipschutz</td>
<td>ANTISOCIAL</td>
<td>0.065</td>
<td>0.067</td>
</tr>
<tr>
<td>Dissertation McCabe</td>
<td>ANTISOCIAL</td>
<td>0.129</td>
<td>0.048</td>
</tr>
<tr>
<td>Dissertation Demondson</td>
<td>ANTISOCIAL</td>
<td>0.394</td>
<td>0.001</td>
</tr>
<tr>
<td>Dissertation Tyas Jones</td>
<td>ANTISOCIAL</td>
<td>0.165</td>
<td>0.008</td>
</tr>
<tr>
<td>Dissertation Brown</td>
<td>ANTISOCIAL</td>
<td>0.110</td>
<td>0.006</td>
</tr>
<tr>
<td>Journal Cooke</td>
<td>ANTISOCIAL</td>
<td>0.113</td>
<td>0.006</td>
</tr>
<tr>
<td>Journal Edwards</td>
<td>ANTISOCIAL</td>
<td>0.463</td>
<td>0.002</td>
</tr>
<tr>
<td>Journal Espelage</td>
<td>ANTISOCIAL</td>
<td>0.050</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal Genesman</td>
<td>ANTISOCIAL</td>
<td>0.120</td>
<td>0.001</td>
</tr>
<tr>
<td>Journal Guim</td>
<td>ANTISOCIAL</td>
<td>0.512</td>
<td>0.012</td>
</tr>
<tr>
<td>Journal Hsu</td>
<td>ANTISOCIAL</td>
<td>0.167</td>
<td>0.002</td>
</tr>
<tr>
<td>Journal McNair</td>
<td>ANTISOCIAL</td>
<td>0.079</td>
<td>0.002</td>
</tr>
<tr>
<td>Journal Ruby</td>
<td>ANTISOCIAL</td>
<td>0.095</td>
<td>0.002</td>
</tr>
<tr>
<td>Journal Scholl</td>
<td>ANTISOCIAL</td>
<td>0.292</td>
<td>0.015</td>
</tr>
<tr>
<td>Journal Taus</td>
<td>ANTISOCIAL</td>
<td>0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>Journal Overall</td>
<td>ANTISOCIAL</td>
<td>-0.060</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Figure 20. Funnel Plot of Studies Reporting Antisocial Outcomes

Figure 21. Forest Plot of Antisocial Outcomes by Report Type
Knowledge

The distribution of the observed effect size values appear biased to the right of the overall standardized difference in means (0.508). By using Duval and Tweedie’s trim and fill to include four hypothetical studies that would balance out the distribution of effect sizes on this outcome, the overall standardized mean effect would be adjusted by -0.386 (0.122). A comparison of the mean effect sizes of published journal articles (0.675) to dissertations (0.056) yields a between-groups Q-statistic of 6.734 (df=1; p=0.010) which is statistically significant. However, the variances of three out of the four imputed studies are less than 0.2, suggesting fairly large scale studies would have to exist with fairly strong negative effect sizes in order to counter the effects of the large scale studies with highly positive effect sizes that were observed.

Figure 22. Funnel Plot of Studies Reporting Knowledge Outcomes
### Table 17. Trim and Fill Adjustments on Studies Reporting Knowledge Outcomes

**Duval and Tweedie’s trim and fill**

<table>
<thead>
<tr>
<th>Group by Report Type</th>
<th>Study name</th>
<th>Outcome</th>
<th>Std diff in means</th>
<th>Std error</th>
<th>Variance</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>Z-Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissertation</td>
<td>Gruber</td>
<td>KNOWLEDGE</td>
<td>0.389</td>
<td>0.078</td>
<td>0.006</td>
<td>0.235</td>
<td>0.543</td>
<td>4.962</td>
<td>0.000</td>
</tr>
<tr>
<td>Dissertation</td>
<td>Lipschutz</td>
<td>KNOWLEDGE</td>
<td>0.490</td>
<td>0.262</td>
<td>0.063</td>
<td>0.033</td>
<td>1.683</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>Dissertation</td>
<td>Rosen</td>
<td>KNOWLEDGE</td>
<td>0.326</td>
<td>0.146</td>
<td>0.023</td>
<td>0.091</td>
<td>0.535</td>
<td>4.747</td>
<td>0.000</td>
</tr>
<tr>
<td>Dissertation</td>
<td>-</td>
<td>-</td>
<td>0.056</td>
<td>0.204</td>
<td>0.042</td>
<td>-0.343</td>
<td>0.371</td>
<td>-4.16</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>Brown</td>
<td>KNOWLEDGE</td>
<td>1.981</td>
<td>0.134</td>
<td>0.018</td>
<td>1.717</td>
<td>2.244</td>
<td>14.727</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>Edwards</td>
<td>KNOWLEDGE</td>
<td>0.056</td>
<td>0.047</td>
<td>0.002</td>
<td>0.024</td>
<td>0.082</td>
<td>5.87</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>Hart</td>
<td>KNOWLEDGE</td>
<td>0.968</td>
<td>0.160</td>
<td>0.032</td>
<td>0.615</td>
<td>1.322</td>
<td>5.374</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>McMahon, 2000</td>
<td>KNOWLEDGE</td>
<td>0.119</td>
<td>0.052</td>
<td>0.000</td>
<td>0.115</td>
<td>0.124</td>
<td>54.493</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>McMahon, 2003</td>
<td>KNOWLEDGE</td>
<td>0.335</td>
<td>0.087</td>
<td>0.000</td>
<td>0.154</td>
<td>0.507</td>
<td>3.834</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>Neace</td>
<td>KNOWLEDGE</td>
<td>1.113</td>
<td>0.069</td>
<td>0.005</td>
<td>0.777</td>
<td>1.820</td>
<td>16.029</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>Sprague</td>
<td>KNOWLEDGE</td>
<td>0.189</td>
<td>0.026</td>
<td>0.001</td>
<td>0.059</td>
<td>0.329</td>
<td>3.732</td>
<td>0.000</td>
</tr>
<tr>
<td>Journal</td>
<td>-</td>
<td>-</td>
<td>0.875</td>
<td>0.123</td>
<td>0.015</td>
<td>0.433</td>
<td>0.346</td>
<td>5.466</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>KNOWLEDGE</td>
<td>0.387</td>
<td>0.068</td>
<td>0.006</td>
<td>-0.318</td>
<td>0.991</td>
<td>1.254</td>
<td>0.210</td>
</tr>
</tbody>
</table>

**Random Effects**

<table>
<thead>
<tr>
<th>Studies</th>
<th>Fixed Effects</th>
<th>Random Effects</th>
<th>Q Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trimmed</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>96</td>
<td>0.12202</td>
<td>0.11775</td>
<td>0.12829</td>
</tr>
</tbody>
</table>

**Figure 23. Forest Plot of Knowledge Outcomes by Report Type**

The analyses for publication bias employed in this meta-analysis suggest that studies reporting prosocial and antisocial outcomes do not significantly vary by report type (i.e., published journal article or dissertation) with regards to the effect sizes reported. Adjusted mean effect sizes after applying Duval and Tweedie’s trim and fill method did not change substantially for prosocial or antisocial outcomes, but the adjusted mean effect for knowledge outcomes did change substantially after applying a trim and fill to the set of studies.
CHAPTER V

DISCUSSION

Summary of Findings

The present study adds to a growing body of evidence concerned with school-based violence prevention. Through the social ecological model of youth violence prevention, it is understood that violence and its prevention are multiply determined. Previous research (Bernat et al., 2012; Herrenkohl et al., 2000) implicated school factors in the prevention of youth violence. A number of these factors were examined in the present study through meta-analysis with regards to Second Step, one of the most prominent and widely recognized SEL programs available. To date, this research represents the only attempt to synthesize the effects of a single school-based SEL program. Compared to other meta-analyses on SEL programs and youth violence prevention, the present study demonstrates weaker positive effects than those conducted with a variety of programs. It is also the only study that presents results on knowledge acquired from a curriculum separately from positive and negative social behaviors. Findings are based on research reports that were identified through a systematic search for published journal articles as well as unpublished dissertations. Eligible reports included unadjusted quantitative data on the program’s effect on students’ prosocial outcomes, antisocial outcomes, and knowledge. Three of the most often cited journal articles (Frey et al., 2005; Orpinas, Parcel, McAlister, & Frankowski, 1995; Van
Schoiack-Edstrom, Frey, & Beland, 2002) about Second Step were not included in the present meta-analysis for a variety of reasons. The study by Frey et al. (2005) reported on a sample of data that was also represented in a dissertation which was also not included because outcome variables of interest were presented only as functions of goal preferences and hostile attribution biases and other covariates in a MANCOVA. The study authored by Orpinas et al. (1995) did not meet inclusion criteria because it only reported baseline adjusted outcome data. The Van Schoiack-Edstrom, Frey, and Beland (2002) study was not included because it only measured the effects of Second Step on students’ attitudes. Of the studies which did meet inclusion criteria, some measured changes in the same sample of students over time before and after participation in Second Step, and other studies compared treatment groups to non-treatment comparison groups.

With respect to these different types of study design, different main effects were found, depending on the outcome measured. Based on single-group repeated-measures studies of Second Step, significant positive effects were demonstrated for prosocial outcomes (ES= 0.09) and knowledge (ES= 0.463). Knowledge was also significantly increased in independent group design studies (ES= 0.767). Participation in Second Step did not appear to significantly impact antisocial outcomes, regardless of the study design. Therefore, the body of evidence is mixed in terms of addressing the primary research question of the present study. The null hypotheses tested in the first research question was that there is no difference between students who participated in the Second Step program and students who have not participated in the program on prosocial, antisocial, or knowledge outcomes. The alternative hypothesis that program participants would experience increased prosocial outcomes, decreased antisocial outcomes, and increased
knowledge is partially supported by the findings. Thus, the null hypothesis that there is no difference may be rejected.

Research questions related to moderator variables were also addressed by the research findings of the present study. However, one variable that was intended to be included in the moderator analyses could not be tested due to an extremely low sample of studies representing different grade level categories. Almost all included studies evaluated the K-5 grade level packages, and it was not possible to draw meaningful comparisons between this grade level package and those developed for preschool students or for middle school students. The results of this analysis would have had practical implications related to the effect of different packages on the age groups for which they are intended. Had this analysis been possible, findings may have informed educators about the age groups most likely to benefit from participation in the program. The developers of Second Step may have also found practical utility in the findings of such an analysis as well. Findings comparing the effects of different grade level packages would be helpful in targeting revision efforts. Although the grade level moderator analysis could not be conducted in the present study, other research questions related to moderating effects were addressed in the present study.

Effects of Second Step appear to be moderated by geographical region, whereby European students demonstrated greater antisociality after intervention than American or Canadian students. This result was based on fourteen studies conducted in USA/Canada and two studies conducted in Germany. A negative program effect was demonstrated by significantly increased antisociality in students after participation in Faustlos, the German adaptation of the Second Step program. This result was not expected, given the
recognition and accolades bestowed upon the original Second Step program. In terms of
the social ecological model of youth violence prevention, this finding could be
interpreted to suggest a greater role for macrosystem influences on antisocial outcomes.
However, due to the severely limited number of studies representing the continent of
Europe, it would be imprudent to make this broad claim based on the data available. A
more plausible explanation for the observed effect may be that the Faustlos program
diverges from the Second Step program in some meaningful way.

Although there was no significant difference between levels of implementation
scale based on the between-groups Q-statistic alone, studies that reported on the
outcomes of medium-scale implementation indicated slight improvements in prosocial
outcomes, whereas high-scale or low-scale implementations did not. Knowledge
increased in students participating in the Second Step program without regard for level of
implementation scale. Effect sizes on antisocial outcomes were non-significant across all
three levels of implementation scale. The finding of non-significant moderating effect by
implementation scale is somewhat unexpected. Both the social ecological model of
youth violence prevention and social learning theory, to which Second Step ascribes,
suggest a higher saturation of a program in the ecological system of a school setting
would contribute to greater program effects. According to these theoretical bases,
relatively high proportions of teachers and students familiar with the concepts taught
through the Second Step curriculum would create a type of synergy in terms of the
positive effects of the program. However, this was not the case based on the data
available. One practical implication of this finding is that educators are afforded a great
deal of flexibility in terms of how large a scale they want to implement Second Step. The
effects of the program on student outcomes are not significantly influenced by implementation scale, and therefore there should be little concern that students would be missing out on some putative program effects related to high saturation. Because Second Step is designed to be delivered as a universal tier 1 intervention, schools or districts may still be interested in providing materials for all students. However, if they are unable to make a large commitment of resources towards the purchasing of the program, there need not be reason for concern.

Related to resource allocation, time, effort, and money spent on training by authorized parties did not significantly improve program outcomes compared to the alternatives. In the studies included in the analysis of the potential moderating effects of training mode, no significant difference in student outcomes was observed for the different modes of implementer training tested. This finding is not expected. Official training materials delivered by authorized parties to teachers was expected to contribute to greater positive effects compared to cases in which trainings were not provided by authorized parties. However, based on the data available, this expectation was not supported. Online training materials are made available to teachers implementing Second Step and included in the purchase price of the kits. On-site in-person trainings may be included as an additional service, and there may be a difference between this type of training and others, but this hypothesis was not tested in the present study.

Dependent variable reporter was the final moderator variable tested in the present study. Based on the data available, student outcomes reported by teachers did not significantly differ from student outcomes based on self-report. The effects of Second Step are observed by teachers and students consistently. This variable was only tested on
prosocial and antisocial outcomes because all knowledge outcomes were measured solely using student sources. The consistency of teacher reports and student reports may be interpreted in a number of ways. One way of interpreting this finding is that the instruments used to measure outcomes are precise and possess good psychometric properties related to convergent validity. Another way of interpreting this finding is that the effects of Second Step are robust and withstand potential differences in perspective based on the reporter’s position within the social ecology of a classroom in which Second Step was implemented.

In summary, only geography appeared to have a significant moderating effect on the impact of Second Step. This could suggest that the effects of Second Step are impervious to factors related to implementation scale, training, or dependent variable reporter. Despite the stability of program effects demonstrated across the potential moderator variables tested, there may be other moderators of effect that were not tested. Based on the $I^2$ values ranging from 35.523% to 98.481% in the main effects analyses, factors responsible for study heterogeneity are posited to exist in addition to the heterogeneity based on sampling error. The variables tested as potential moderators in the present study do not appear to account for this heterogeneity. Other moderating variables may be partially responsible for this heterogeneity. A likely candidate may be differences in student risk for aggression prior to intervention. Hussey and Flannery’s study (2007) indicated significant differences between high baseline and low baseline students on measures of aggression. Similarly, Frey and colleagues (2005) reported low baseline scores in antisocial behavior showed no change, whereas control students increased. In the study conducted by Orpinas and colleagues (1995), evidence exists to
support the notion that Second Step had a greater effect on higher risk students. Because this predisposition was treated as a covariate in these studies, and because covariate-adjusted data was not included in the present meta-analysis, this information could not be included in the synthesis. Furthermore, participant risk for violence or aggression was not reported in any other studies meeting inclusion criteria.

Although the evidence bears out non-significant moderation of program effects based on most of the variables tested, it should be noted that significant program effects were observed for a number of the study subgroups. On prosocial outcomes USA/Canada studies had a significant positive effect (ES= 0.081). On antisocial outcomes, European studies had a significant negative effect (ES= -0.363). Medium-scale implementation had a significant positive effect on prosocial outcomes (ES= 0.156), and high-scale implementation had a significant positive effect on knowledge outcomes (ES= 0.488). Informal/not reported training appeared to have a significant positive effect on prosocial outcomes (ES=0.097) and knowledge outcomes (ES=0.696). Self-reported outcomes appeared to have a significant negative effect on antisocial outcomes (ES= -0.117). Among all of these the European studies measuring antisocial outcomes is the only subgroup that was significantly different than its comparison subgroup(s).

The estimates of heterogeneity produced by this sample of studies ought to be interpreted with caution due to small cell size. In some cases only two or three studies represented a level of a moderator. The tau-squared statistic is central in describing the heterogeneity and potential moderator effects in a meta-analysis. Borenstein (2009) described the difficulties of estimating tau-squared from a small number of studies. The
findings of the present study are likely subject to the uncertainty related to the dynamics described by Borenstein.

**Implications**

Overall, it appears that Second Step, as a universal Tier 1 program, does not significantly affect prosocial or antisocial outcomes in students as rated by themselves or by teachers. However, student knowledge of the curriculum increased as a result of participating in the program. Because this increase in knowledge does not appear to be accompanied by decreased antisocial behaviors or increased prosocial behaviors, a number of questions are raised about the role of Second Step and of similar programs in school-based violence prevention efforts. Is the role of a violence prevention of social emotional learning curriculum to merely provide a foundation of knowledge? If so, Second Step ought to be considered as a highly effective program. However, if the intended role of Second Step and programs like it are to produce substantive and observable changes in student behavior, the research fails to support this function. However, some of the limitations of the primary research studies on Second Step may have some impact on this conclusion. For instance, the durations of studies included in this meta-analysis may have been too short to observe any effects this increased knowledge may have had on behavioral outcomes over a longer course of time. In a universal prevention program like Second Step, the amount of time students have to apply and internalize knowledge gained from programs may be of greater importance than implementation scale or the type of training received by teachers. However, what is unknown from the primary studies was the amount of time students might have been exposed to the program prior to the research study. Whether or not increased knowledge
from the curriculum translates into improved behaviors, educators may be interested in appealing to students’ increased knowledge and reminding them of lessons learned as conflicts arise throughout the school year. However, other meta-analyses of SEL programs and interventions intended to reduce aggression and disruptive behaviors did, in fact, find more positive findings. These positive findings were synthesized from studies that may have also been characterized by similar time constraints as the studies included in the present meta-analysis. These observations aside, educators can use the common language and framework espoused in Second Step to guide students in managing their emotions and resolving conflict in non-aggressive ways.

The present study was unable to replicate results from previous meta-analyses on SEL programs and school-based interventions for aggression that demonstrated positive program effects by either reducing aggression or improving prosocial behavior. These findings are surprising considering the accolades and recognition the program has garnered and the prominence of Second Step among SEL programs available to educators. As one example of Second Step’s prominence, it has been recognized as a SELECT Program by the Collaborative for Academic, Social, and Emotional Learning (CASEL). Second Step does meet CASEL’s criteria for recognition in that it is well-designed, it delivers training and implementation supports, and its positive impacts have been documented in at least one carefully conducted evaluation. CASEL, a widely-recognized authority on SEL, has set a relatively low standard for evidence of effectiveness. The present study synthesized the results of 24 carefully conducted evaluations, and the evidence of Second Step’s impact is mixed.
The results suggest that universal prevention programs such as Second Step play a small role in the prevention of youth violence. Programs like Second Step may be effective in imparting knowledge about concepts relevant to violence prevention. The results of the present study suggest educators who are interested in preventing violence ought to consider providing other supports to students in addition to packaged curricula like Second Step. Second Step may play a vital role as one component of a multi-tiered system of supports for students in schools who present any level of risk for violence. Educators ought to consider other universal tier 1 supports for students in addition to imparting knowledge through programs like Second Step. Implementation of programs like Second Step may be couched within a system of positive behavior supports. In addition to other tier 1 supports, targeted interventions for students presenting elevated risk for violence or aggression should also be in place.

**Implications on Social Ecological Model of Youth Violence Prevention**

One objective at the outset of this research project was to evaluate the social ecological model of youth violence prevention through empirical data. The strategy to isolate a single program as an input variable remains a good one. Though not enough, studies were found eligible for inclusion to take advantage of the potential strengths of this approach. However, with what was found about implementation scale, we have to suspend drawing any conclusions about empirical support of the model from this study. If the findings had found higher saturation leading to better student outcomes, this study would have provided empirical support for the theoretical proposition that direct and indirect influences from multiple levels of the ecological model converge to impact outcomes related to preventing youth violence. However, simply because that finding
was not borne out does not mean that the model should be abandoned. Decisions about
the utility of this model should not be based on one study alone. The social ecological
model remains widely held as a valid and useful model for conceptualizing youth
violence. Researchers should be encouraged to undertake continued efforts to test the
model as it pertains to youth violence. The design and findings of the present study may
be used as a starting point for other researchers to evaluate the social ecological model of
youth violence prevention.

Limitations

One of the most common limitations in meta-analysis is want for a greater corpus
of primary literature. This need is underscored in cases in which multiple study designs,
multiple dependent variables, and multiple moderator variables are represented in the
literature. It is uncertain whether the statistical tests used in the moderator analyses in the
present study had sufficient power, as no a priori power analysis was conducted. Another
common limitation related to the extant primary literature is the absence of certain data
necessary for effect size calculation. As stated earlier in this report pre/post correlation
coefficients were absent from eight primary studies (Beisly, 2011; Brown, Jimerson,
Dowdy, Gonzalez, & Stewart, 2012; Hart, Dowdy, Eklund, Renschaw, Jimerson, Jones, &
Earhart, 2009; McCabe, 2000; McMahon & Washburn, 2003; Neace & Munoz, 2012;
Rosen, 2013; Schick & Cierpka, 2005) and imputed with a fixed value of 0.500. While
this correlation coefficient provides an estimate of program effect in those samples,
imputation of another value may produce a different estimate of effect size. Most studies
included in the present meta-analysis were conducted in the United States. Whereas the
objective was to collect studies representative of Second Step’s wide dissemination in
other nations around the world, these studies could not be located.

In studies reporting prosocial outcomes, there was evidence of some positive
reporting bias in journal articles compared to dissertations. Similarly, published journal
articles reporting antisocial outcomes presented a less negative effect than dissertations.
However, the differences between journal articles and dissertations were not statistically
significant. On the other hand, there is stronger evidence for possible publication bias in
studies reporting knowledge outcomes based on the statistically significant difference
between the effect sizes reported in journal articles compared to dissertations.

**Directions for Future Research**

Although there appeared to be little moderation of SS ES using the categorical
moderators, a follow-up analysis using meta-regression may highlight relationships
among these moderators that may add to the understanding of programs like Second Step
within a social ecological model of youth violence prevention. The dimension of time is
very salient to prevention research. In order to better examine the role of SEL programs
in youth violence prevention, future studies ought to incorporate longitudinal designs
measuring multiple outcomes. Future studies of Second Step and similar programs
should include pre-intervention baseline data on student risk for aggression or violence.
Data related to knowledge acquired should be measured immediately after intervention.
Data on knowledge maintenance at a follow-up period ought to be collected as well as
data measuring antisocial and prosocial outcomes at follow-up. More research is needed
on the implementation contexts of SEL programs. Many studies report population
statistics on student ethnicity and free/reduced lunch status, but the relevance of this
information is seldom discussed in terms of moderation of program effectiveness. A number of multi-site studies have already been conducted on Second Step. There is a need for future multi-site studies in which student characteristics and contextual variables are carefully measured and considered in the design of the studies. Implementation contexts may impact the effects of programs such as Second Step, and these contexts ought to serve as the focal point of future studies on Second Step and similar programs.
APPENDIX A

INCLUDED STUDIES
<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Participants</th>
<th>Outcomes Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beisly, 2003</td>
<td>62</td>
<td>prosocial</td>
</tr>
<tr>
<td>Bogue, 2011</td>
<td>44</td>
<td>antisocial</td>
</tr>
<tr>
<td>Brown, 2012</td>
<td>165</td>
<td>antisocial</td>
</tr>
<tr>
<td>Cooke, 2007</td>
<td>639</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Edwards, 2005</td>
<td>318</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Espelage, 2013</td>
<td>3,616</td>
<td>antisocial</td>
</tr>
<tr>
<td>Grossman, 1997</td>
<td>790</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Gruber, 2007</td>
<td>175</td>
<td>knowledge</td>
</tr>
<tr>
<td>Grumm, 2013</td>
<td>155</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Hart, 2009</td>
<td>149</td>
<td>knowledge</td>
</tr>
<tr>
<td>Holsen, 2008</td>
<td>743</td>
<td>prosocial</td>
</tr>
<tr>
<td>Hussey, 2007</td>
<td>239</td>
<td>antisocial</td>
</tr>
<tr>
<td>Lipschutz, 2010</td>
<td>66</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>McCabe, 2000</td>
<td>83</td>
<td>antisocial</td>
</tr>
<tr>
<td>McMahon, 2000</td>
<td>42</td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>McMahon, 2003</td>
<td>149</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Neace, 2012</td>
<td>388</td>
<td>knowledge</td>
</tr>
<tr>
<td>Osmondson, 2013</td>
<td>656</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Rosen, 2013</td>
<td>61</td>
<td>knowledge</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Ruby, 2010</td>
<td>944</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Schick, 2005</td>
<td>718</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Sprague, 2001</td>
<td>3,699</td>
<td>knowledge</td>
</tr>
<tr>
<td>Taub, 2001</td>
<td>70</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
<tr>
<td>Tynes-Jones, 2006</td>
<td>125</td>
<td>antisocial</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prosocial</td>
</tr>
</tbody>
</table>
APPENDIX B

CODING PROTOCOL
SECOND STEP PROGRAM EFFECTIVENESS
META-ANALYSIS DATA CODING INSTRUMENT

SECTION A
BIBLIOGRAPHICAL INFORMATION AND SCREENING

A1. Study ID# __ __ __ [STUDYID]

A2. Coding Date _Y_ _Y_ - _M_ _M_ - _D_ _D_ [CODDATE]

A3. Coder initials __ __ __ [CODER]

A4. Primary author (LN, FI) [AUTHOR]

______________________________________________

A5. Year of publication __ __ __ __ [PUBYR]

A6. Bibliographic info in APA format: [REF]

______________________________________________

______________________________________________

A7. Does study report student outcomes of implementing the Second Step
program or one of its translated adaptations (e.g. Steg for Steg,
Faustlos, Paso Adelante, etc) ? [OC]
A8. Indicate the type of paper/study below:

☐ 1. outcome/program/intervention evaluation (CONTINUE)
☐ 2. review of social competence outcome studies (STOP)
☐ 3. position paper, editorial, book review (STOP)
☐ 4. guidelines for treatment or intervention (STOP)
☐ 5. qualitative research (STOP)
☐ 98. other: _________________________________ (STOP)
☐ 99. cannot tell (STOP)

A9. Indicate the source of the paper below:

☐ 1. peer-reviewed journal
☐ 2. Dissertation or thesis
☐ 3. technical report
☐ 98. other:
   Specify _________________________________
☐ 99. cannot tell

A10. Indicate the type of source utilized to access the publication.

[DTBASE]

☐ 1. electronic database
   Specify_______________________________
☐ 2. electronic book search
☐ 3. web search
   Insert URL:
Specify________________________________________

4. reference in a book or study
Specify________________________________________

5. peer or expert
Specify________________________________________

98. other
Specify________________________________________

99. cannot tell

A11. Type of design
[DESIGN]

1. Randomized Controlled Experiment
2. Quasi-Experiment With No Treatment Control Group
3. Quasi-Experiment With Alternate Treatment Control Group
4. Single Group Pretest-Posttest Design
98. Other
Specify________________________________________

99. cannot determine

Final Decision regarding this study

A12. Should this study be retained for further analysis?
[INCLUDE]

1. yes
2. no
99. unsure based upon information obtained up to this point
READ FIRST**

**Sometimes the scale of the implementation and the scale of the evaluation study are different. For example, the intervention can be implemented in 5 schools in a 10-school district, but researchers may only use 1 treatment school and 1 control school in the evaluation study. Items B1-B8 address the potential for this type of dynamic to occur in the literature.

B1. Indicate the total number of students included in the IMPLEMENTATION of Second Step. *(Do not count CXN if Second Step wasn’t implemented with CX. Often same as TXN, except in cases in which more students received intervention than the subset who participated in the evaluation.)*

[IMPN]

B2. Indicate the number of classrooms included in the IMPLEMENTATION of Second Step.

[IMPNCLASS]

Indicate number:________________________________

(1= one classroom or small group; 99= cannot tell)

B3. Indicate the number of schools included in the IMPLEMENTATION of Second Step.
Indicate number: ____________________________________
(1 = within one school; 99 = cannot tell)

B4. Indicate the number of districts included in the IMPLEMENTATION of Second Step.

Indicate number: ____________________________________
(1 = within one district; 99 = cannot tell)

B5. Indicate the number of classrooms included in the COMPARISON group of the EVALUATION of Second Step.

Indicate number: ____________________________________
(1 = one classroom or small group; 99 = cannot tell)

B6. Indicate the number of classrooms included in the TREATMENT group of the EVALUATION of Second Step.

Indicate number: ____________________________________
(1 = one classroom or small group; 99 = cannot tell)

B7. Indicate the number of schools included in the COMPARISON group of the EVALUATION of Second Step.

Indicate number: ____________________________________
(1 = within one school; 99 = cannot tell)
B8. Indicate the number of schools included in the TREATMENT group of the EVALUATION of Second Step. [TXEVNSCHOOL]

Indicate number:______________________________
(l= within one school; 99= cannot tell)

B9. Indicate the number of districts included in the COMPARISON GROUP of the EVALUATION of Second Step. [CXEVNDIST]

Indicate number:______________________________
(l= within one district; 99= cannot tell)

B10. Indicate the number of districts included in the TREATMENT GROUP of the EVALUATION of Second Step. [TXEVNDIST]

Indicate number:______________________________
(l= within one district; 99= cannot tell)

B11. Indicate the grade levels of the classrooms included in the IMPLEMENTATION of Second Step. [IMPgrade]

☐ 1. Early childhood/pre-K
☐ 2. Elementary school (K-5)
☐ 3. Middle school/Junior High (6-8)
☐ 4. Combination of 1 & 2
☐ 5. Combination of 1 & 3
☐ 6. Combination of 2 & 3
☐ 7. Combination of 1, 2, & 3
B12. Indicate the grade levels of the classrooms included in the COMPARISON group of the EVALUATION of Second Step.

[CXEVGRADE]

☐ 1. Early childhood/pre-K
☐ 2. Elementary school (K-5)
☐ 3. Middle school/Junior High (6-8)
☐ 4. Combination of 1 & 2
☐ 5. Combination of 1 & 3
☐ 6. Combination of 2 & 3
☐ 7. Combination of 1,2,& 3
☐ 98. Not school-based (specify)______________________________
☐ 99. cannot tell

B13. Indicate the grade levels of the classrooms included in the TREATMENT group of the EVALUATION of Second Step.

[TXEVGRADE]

☐ 1. Early childhood/pre-K
☐ 2. Elementary school (K-5)
☐ 3. Middle school/Junior High (6-8)
☐ 4. Combination of 1 & 2
☐ 5. Combination of 1 & 3
☐ 6. Combination of 2 & 3
☐ 7. Combination of 1,2,& 3
☐ 98. Not school-based (specify)______________________________
☐ 99. cannot tell
B14. Was Second Step implemented as a school-wide intervention in this study?  

[IMPSCHWIDE]

☐ 1. Yes, at least one entire school participated in the implementation of Second Step.

☐ 2. No, but more than half of the classrooms in a school participated in the implementation of Second Step.

☐ 3. No, less than half of the classrooms in a school participated in the implementation of Second Step.

☐ 98. Other (specify) ________________________________

☐ 99. Cannot tell

B15. Was Second Step implemented as a district-wide intervention in this study?  

[IMPDISWIDE]

☐ 1. Yes, at least one entire district participated in the implementation of Second Step.

☐ 2. No, but more than half of the schools in a district participated in the implementation of Second Step.

☐ 3. No, less than half of the schools in a district participated in the implementation of Second Step.

☐ 98. Other (specify) ________________________________

☐ 99. Cannot tell

B16. Briefly summarize how schools or classrooms were selected for inclusion in the EVALUATION of Second Step (e.g. random assignment, random selection, matching, cannot tell) and the page number where this information can be found:  

[EVSELECT]
SECTION C

Intervention Implementation

C1. If stated, what is the primary intention of implementing Second Step in the study?

[GOAL]

☐ 1. Enhancing SEL is explicitly stated as the primary goal
☐ 2. Violence prevention is explicitly stated as the primary goal
☐ 3. Increased positive outcomes (e.g. building social skills, building positive relationships, improving school climate) are stated as the primary goal (please specify) __________________________

☐ 4. Decreased negative outcomes (e.g. less conflict, less disciplinary infractions) are stated as the primary goal (please specify) __________________________________________

☐ 5. Combination of the above choices
☐ 99. cannot tell

C2. How was Second Step implemented? [SOLO]

☐ 1. As the sole intervention and focus of the study
☐ 2. As one component of a host of simultaneously implemented interventions ALSO evaluated in the study
☐ 3. As one component of a host of simultaneously implemented interventions and the ONLY intervention evaluated among them
☐ 99. cannot tell
C3. Indicate whether screening procedures were used to determine participation in Second Step?

[SCREEN]

- 1. Yes, screening procedures were used
- 2. No, screening procedures were not used
- 99. Cannot tell

C4. At what tier of service delivery was Second Step implemented?

[TIER]

- 1. Tier 1 / Universal
- 2. Tier 2 / Selected
- 3. Tier 3 / Indicated
- 4. Combination of 1,2, or 3
- 99. cannot tell

C5. Implementation location

[METRO]

- 1. urban
- 2. suburban
- 3. rural
- 4. more than one of the above within one geographic locale
- 5. More than one of the above across multiple geographic locales
- 99. cannot tell

C6. Indicate the geographic location of the implementation of Second Step.

[GEO]

- 1. USA/Canada
2. Latin American nation (e.g. Mexico, Chile, Guatemala, Brazil)
   Specify__________________________________
3. European nation (e.g. Norway, Great Britain, Germany)
   Specify__________________________________
4. Asian nation (e.g. Kurdistan, Japan, Laos, India)
   Specify__________________________________
5. African nation (e.g. Morocco, Nigeria, South Africa, Ethiopia)
   Specify__________________________________
6. Australasian nation/region (e.g. Australia, New Zealand, Fiji)
   Specify__________________________________
7. Other
   Specify__________________________________
99. cannot tell

C7. Who delivered the intervention?

1. Teacher
2. Clinician
3. Researcher
4. Combination of 1&2
5. Combination of 1&3
6. Combination of 2&3
98. Other
   Specify _________________________________
99. cannot tell
C8. Was the implementation of the program monitored by the researcher or program personnel to assess whether it was delivered as intended?

[FIDMON]

☐ 1. yes
☐ 2. no
☐ 99. cannot tell

C9. To what extent were the school-based components of the program delivered with fidelity?

[FIDOK]

☐ 1. Covered all lessons
☐ 2. Covered at least 90% of lessons, or at least 90% of teachers report high fidelity
☐ 3. Covered at least 75% of lessons, or at least 75% of teachers report high fidelity
☐ 4. Covered at least 50% of lessons, or at least 50% of teachers report high fidelity
☐ 5. Covered LESS than 50% of lessons, or FEWER than 50% of teachers report high fidelity
☐ 99. cannot tell

C10. To what extent were the home-based components of the program delivered with fidelity?

[FIDHOME]

☐ 1. Home-based components were disseminated for at least half of the lessons
☐ 2. Home-based components were disseminated for less than half of the lessons
3. Home-based components were not disseminated at all

99. cannot tell

C11. Duration of intervention

Enter the actual maximum duration of the intervention implementation in number of weeks

_______

99. cannot determine

C12. Indicate the level of training received by implementers.

1. online training using official materials
2. on site training by authorized party
3. combination of 1 & 2
4. no formal training with official materials/personnel

98. other ______

99. cannot tell
SECTION D  
Participants  

Categories of participant descriptions shall be coded for treatment (TX) and comparison or control (CX) groups. In many instances, these characteristics are reported in the aggregate. In those cases, simply enter the same value for CX and TX.

D1. Indicate the PREDOMINANT level of “risk” of juveniles in this group at onset of the study. *Most will be universal UNLESS Second Step was explicitly and specifically targeted towards a selected or indicated group

[CX/TX RISK]

☐ 1. Universal: Normal children, general population, school-wide samples, etc.

Selected: Selected populations are those exhibiting a risk factor for aggression, violence, or related antisocial behaviors.

☐ 2. Selected based on neighborhood, environment, or group characteristics (e.g., inner city, low SES area)

☐ 3. Selected based on individual characteristics (e.g., low reading ability, temperament)

☐ 4. Indicated: Indicated samples are those chosen for intervention because they are displaying aggression, violence, or related antisocial behaviors.

☐ 5. Mixed

☐ 99. Cannot tell
D2. Does the history of the juveniles in this group include aggression, violence, fighting, bullying, assaults, or similar person-directed antisocial behavior, whether officially recorded or not?

[CX/TX RISKHIST]

☐ 1. no. Select this option only if the report(s) clearly indicate that the group has no such history; do not make assumptions.

☐ 2. yes, some juveniles (<50%)

☐ 3. yes, most juveniles (= or >50%)

☐ 4. yes, all juveniles (>95%)

☐ 5. some, but cannot estimate percent

☐ 99. cannot tell

D3. Indicate PREDOMINANT Race/ethnicity of participants (50% or greater to qualify as predominant group) [CX/TX RACE]

☐ 1. Caucasian

☐ 2. African American

☐ 3. Hispanic/Latino

☐ 4. Asian

☐ 98. Other

Specify ________________________________

☐ 99. cannot determine
D4. Indicated socioeconomic status of majority of participants. **If free/reduced lunch is provided, insert proportion as a decimal in a comment.

[CX/TX SESCAT]

- 1. Low (at or below poverty line)
- 2. Working or lower middle class
- 3. Middle class or above
- 4. Combination
- 99. cannot tell

D5. Indicated participant disability

[CX/TX DISAB]

- 1. No disability indicated
- 2. Conduct disorder/ oppositional defiant disorder
- 3. Mood disorder
- 4. Attention deficit-hyperactivity disorder
- 5. Learning disability
- 6. Combination
- 99. cannot tell

D6. Enter the AVERAGE age of the participants in years at the beginning of the study.

[CX/TX AVGAGE]

___________

D7. Enter the age in years of the youngest participants at the beginning of the study.

[CX/TX AGELO]

___________
D8. Enter the age in years of the oldest participants at the beginning of the study. [CX/TX AGEHI]

D9. What was the lowest grade level of the study sample at the beginning of the study. [CX/TX GRADELO]

- 1. Pre-K
- 2. Kindergarten
- 3. 1st grade
- 4. 2nd grade
- 5. 3rd grade
- 6. 4th grade
- 7. 5th grade
- 8. 6th grade
- 9. 7th grade
- 10. 8th grade
- 99. cannot tell

D10. What was the highest grade level of the study sample at the beginning of the study. [CX/TX GRADEHI]

- 1. Pre-K
- 2. Kindergarten
- 3. 1st grade
- 4. 2nd grade
- 5. 3rd grade
- 6. 4th grade
- 7. 5th grade
- 8. 6th grade
9. 7th grade
10. 8th grade
99. cannot tell

D11. How did researchers aggregate participant outcome data?

[DVBREAKOUT]

☐ 1. For each dependent variable, outcome data from participants were reported in the aggregate
☐ 2. For each dependent variable, outcome data was disaggregated by age/grade level
☐ 3. For each dependent variable, outcome data was disaggregated by sex/gender
☐ 4. For each dependent variable, outcome data was disaggregated by both age/grade and by sex/gender
☐ 98. For each dependent variable, outcome data was disaggregated by another variable (specify): ______________________________
SECTION E

DV Dependent Variable Characteristics

One SECTION E should be completed for each dependent variable.

E1. Study ID: Type in the appropriate Study ID [STUDYID]

E2. Identify the DV number per study [DVID]
If there is only one relevant DV per study, enter “1”. Each additional DV in this study should be labeled 2, 3, 4 etc. If there were reported breakouts, each respective DV breakout (i.e. by participant age, grade, gender, etc) receives its own DVID. If there are multiple sources of data (teacher report, self report, parent report, etc), each respective DV source receives its own DVID.

E3. Construct measured, including distinguishing breakout/DV source descriptor (e.g. if the study breaks out by gender, and the construct is aggression, type in “Aggression x Boys” for one DVID and “Aggression x Girls” for the other DVID. There is a separate code for DV sources [E7], so you do not need to put that here in the DVNAME).

[DVNAME]
E4. Type of dependent variable  [DVTYPE]

☐ 1. Physical violence/aggression
☐ 2. Verbal aggression
☐ 3. Aggression: combined or not otherwise specified
☐ 4. Other antisocial behavior
   Specify_______________________________
☐ 5. Positive social behavior
   Specify_______________________________
☐ 6. Knowledge or skills
   Specify_______________________________
☐ 7. Attitudes
   Specify_______________________________
☐ 98. Other
   Specify_______________________________
☐ 99. Cannot tell

E5. Type of measure operationalizing DV  [DVMEASURE]

☐ 1. direct observation
☐ 2. Rating scale/checklist/survey/multi-item measure (e.g. CBCL, etc)
☐ 3. Sociometric
☐ 4. School records/office disciplinary referrals
   5. SECOND STEP proprietary assessment
☐ 98. Other: _______________________________
☐ 99. cannot determine or not reported
E6. Origin of measure

☐ 1. Pre-existing measure
☐ 2. Measure was developed for this study

E7. Respondent or source of data

☐ 1. Parent or caregiver report
☐ 2. Teacher/school professional report
☐ 3. Independent observer
☐ 4. Self-report
☐ 5. Peer
☐ 6. Multiple sources
☐ 99. cannot determine or not reported

E8. Do higher values indicate greater desired behaviors/skills?

☐ 1. yes
☐ 2. no, it is meant to indicate higher undesired behaviors or symptoms

E9. Enter Reliability Coefficient (if available).

Use two digits and a decimal point, e.g., .96. You may use any type of reliability coefficient (test-retest, Cronbach’s alpha, etc.) and any sample. That is, if the researchers provide a reliability coefficient from another study, you may use it here.

____________
E10. If you entered a reliability coefficient, indicate the type of coefficient you entered. [RELTYPE]

If the study reports more than one type of coefficient, select only one in order of priority from 1 to 4, according to the list below.

- 1. internal consistency (e.g., split half, Cronbach’s alpha or alpha-reliability, Kuder-Richardson reliability, etc.)
- 2. test-retest reliability (e.g., test-retest reliability, coefficient of stability)
- 3. inter-rater reliability (e.g., interrater reliability, percent agreement, Kappa coefficient)
- 4. alternate form reliability (e.g., coefficient of equivalence)

E11. Source of the reliability coefficient. [RELSOURCE]

Indicate whether the reliability coefficient you entered above was derived from the current sample or some other group of individuals (e.g., sometimes author(s) will provide reliability coefficients given by the developers of the instrument).

- 1. all or part of the sample of individuals from the study you are coding
- 2. the instrument (e.g., test manual, other studies by the test developer); this implies that the sample of individuals upon which the reliability was determined is NOT the sample of individuals from the study you are coding
3. studies by other researchers (but not the test developer); this implies that the sample of individuals upon which the reliability was determined is NOT the sample of individuals from the study you are coding

99. cannot tell

E12. Reliability proxy

Use the available information to assess the approximate reliability of the measure.

1. single item measure (or one observer)
2. multiple item measure with 5 or fewer items (or two observers)
3. multiple item measure with more than 5 items (or more than two observers)
98. Other
   Specify ________

E13. Was data collected regarding maintenance of treatment effects over time (follow-up)?

1. yes (proceed to next item)
2. no
99. cannot determine or unclear

E14. How much time (in months) passed between the end of the study and the collection of follow-up data?

______________________

99. cannot determine or not applicable
SECTION F

Effect Size Data

One SECTION F should be completed for each dependent variable.

F1. Study ID: Type in the appropriate Study ID [STUDYID]
F2. DV ID: Type in the appropriate DV ID [DVID]

F3. Effect size ID. [ESID]

Use this field to number the effect sizes for THIS study. Thus, a study with 10 effect sizes would have the numbers 1 through 10. Start over with 1 for each new study that you are coding.

F4. Page number for this effect size. [PGNUM]

Indicate the page number of the report identified above on which you found the effect size data. If you used data from two different pages, you can type in both, but use a comma or dash between the page numbers.

F5. Type of effect size [ESTYPE]

There are 4 types of effect sizes that can be coded: pretest, posttest, follow-up, and group equivalence (or pretreatment similarity) effect sizes. They are defined as follows:

• Pretest effect size. This effect size measures the difference between a treatment and comparison group before treatment (or at the beginning
of treatment) on the same variable used as an outcome measure, e.g., aggressive behaviors measured before the treatment begins are used as a “pretest” for aggressive behaviors measured after the treatment ends.

• Posttest effect size. This effect size measures the difference between a treatment and comparison group after treatment on some outcome variable. A posttest can occur right after treatment ends or after some delay, but it is distinguished from a follow-up (see below) because it is the first measure taken after treatment ends, regardless of the time period between the end of treatment and posttest measurement.

• Follow-up effect size. Follow-up effect sizes measure the differences between a treatment and comparison group after treatment (as with the posttest effect sizes above), but they involve later measurement waves. That is, some studies may measure the differences between treatment and comparison groups directly after treatment and then 6 months later. The measurement taken at 6 months would be coded as a follow-up effect size.

• Group equivalence effect size. Group equivalence effect sizes are used to code the equivalence of a treatment and comparison groups prior to treatment delivery on variables that might be related to outcome, such as gender, age, ethnicity, and the like. A pretest that is used later in the study as a posttest would not be coded here – you would code it as a pretest effect size. You will ordinarily calculate group equivalence effect sizes as part of the process for the header coding
sheet, rather than as part of the process for the effect size coding sheet.

Type of effect size:

☐ 1. Pretest (for treatment-control comparison on a dependent variable)

☐ 2. Posttest (for treatment-control comparison on a dependent variable)

☐ 3. Follow-up (for treatment-control comparison on a dependent variable)

☐ 4. Group Equivalence (for pretest treatment-control comparisons on variables other than the dependent variables)

It is now time to identify the data you will use to calculate the effect size, and to calculate the effect size yourself if necessary (see below).

Effect sizes can be calculated ONLY from data based on the number of subjects, e.g., mean number of aggressive acts per subject (and the corresponding standard deviation) or proportion of subjects who acted aggressively during a given time period. Effect sizes can NOT be calculated from data based solely on the incidence of events, e.g., total number of aggressive acts per group. Effect sizes can be calculated from subject-based data in a variety of forms; to determine which data you should use for effect size calculation, please refer to the following guidelines:

1. Compute ES from descriptive statistics if possible (means, sds, frequencies, proportions).
2. If adequate descriptive statistics are unavailable, compute ES from significant test statistics if possible (t, F, Chi square, etc.).

3. If significance tests statistics are unavailable or unusable but p value and degrees of freedom (df) are available, determine corresponding t value and compute ES as if t-test had been used.

F6. Which group is favored? [CXORTX]

For treatment-control comparisons, the treatment group is favored when it does "better" than the control group. The control group is favored when it does "better" than the treatment group. Remember that you cannot rely on simple numerical values to determine which group is better off. For example, a researcher might assess the amount of violent behavior, and report this violent behavior in terms of the number of violent acts per subject per day. Less violent behavior is better than more, so in this case a lower number, rather than a higher one, indicates a more favorable outcome.

Sometimes it may be difficult to tell which group is better off, because some studies use surveys or paper and-pencil measures in which it is unclear whether a high score or a low score is more favorable. In these situations, a thorough reading of the text from the results and discussions sections usually can bring to light the direction of effect – e.g., the authors will often state verbally which group did better on the measure you are coding, even when its not clear in the data table. Note that if you cannot determine which group has done better, you will not be able to calculate a numeric effect size. (You will still be able to create an effect size record—just not a numeric effect size.) Remember that every study must produce at least one numeric effect size to be eligible for coding; if you find that you cannot determine which
group has done better for any of the potential effect sizes in a study, the study is not eligible.

F6. Select the group that has done “better”:  
[CXORTX]

- 1. Treatment
- 2. Control
- 3. Neither, Exactly Equal
- 99. Cannot tell

F7. Effect size derived from what type of statistics?  
[STATTYPE]

- 1. N successful (frequencies)
- 2. Proportion successful (percentage successful or not successful)
- 3. Multi-category (polychotomous) frequency or %
- 4. Means and SDs, means and variances, means and standard errors
- 5. Independent T-test
- 6. Dependent T-test
- 7. Probability With N/degrees of freedom
- 8. One-way ANOVA (2 groups, 1 degree of freedom)
- 9. One-way ANOVA (>2 groups, >1 degree of freedom)
- 10. Factorial Design (Repeated measures ANOVA, 2x2 ANOVA, MANOVA, etc.)
- 11. Covariance Adjusted (ANCOVA)
- 12. Chi-square statistic (1 degree of freedom)
- 13. Chi-square (> 2x2 table)
- 14. Nonparametric statistics (Mann Whitney, etc.)
- 15. Correlation coefficient (zero-order)
- 16. Multiple regression
17. Effect sizes

F8. For this effect size, did you use adjusted data (e.g., covariate adjusted means) or unadjusted data? [ADJDATA]

If both unadjusted and adjusted data are presented, you should use the adjusted data. Adjusted data are most frequently presented as part of an analysis of covariance (ANCOVA). The covariate is often either the pretest or some personal characteristic such as socioeconomic status.

1. Unadjusted data
2. Pretest adjusted data
3. Data adjusted on some variable other than the pretest (e.g., socioeconomic status, IQ)
4. Data adjusted on pretest and other variables

F9. Significance information for this comparison. [SIG]

For treatment-control comparisons: Did the authors make any comment about the statistical significance of the difference between the values (e.g., mean test scores) for the two groups you selected, with regard to the dependent variable you have selected, at the time point you have selected for this comparison? Sometimes authors will state that a particular comparison was not significant, but not provide any calculable effect size data. In these cases, you should select "5" for this item. The effect size field should remain blank. In other cases, authors will state that a particular comparison was significant,
but not provide any calculable effect size data. In these cases, you should select “4” for this item. Again, the effect size field should remain blank.

NOTE: the last three options (4, 5, and 6) are for cases for which you have direction (i.e., you know which group is favored) for no effect size information.

☐ 1. Significant result, ES data below
☐ 2. Non-significant result, ES data below
☐ 3. Significance not reported, ES data below
☐ 4. Significant result, no ES data
☐ 5. Non-significant result, no ES data
☐ 6. Significance not reported, no ES data

F10. Variance control techniques. [VARTYPE]

Type of statistical test done for this comparison, if any. The issue here is whether the author(s) used a variance-control technique when analyzing the comparison for which you are calculating an effect size.

☐ 1. No Test
☐ 2. No Report. Use this option when you have significance info, but don’t know the kind of test used.
☐ 3. No variance control techniques (e.g., t-test, oneway ANOVA, z-test, Η2, non-parametric, raw means, etc.)
☐ 4. Variance control techniques used (e.g., ANCOVA, multiple regression, repeated measures ANOVA, adjusted means, etc.)
DATA ENTRY FIELDS FOR EFFECT SIZE CALCULATION

Assigned and Observed N

Assigned N, Observed N. These fields refer to the number of subjects who were originally assigned to the two groups (Assigned N) and to the number of subjects who were actually “observed” or “measured” (Observed N). If you cannot tell how many subjects were originally assigned to a group, take a look at the number of subjects (Observed N) at pretest; you can frequently use pretest sample sizes for assigned N. However, in cases where the authors have removed the subjects who do not have both pretest and posttest measures (such that the pretest N and the posttest N are the same), do not assume that the number of subjects at pretest is the correct number for Assigned N, and leave this field blank. In cases where there is no attrition, the Assigned N is the same as the Observed N. Only use the same numbers for Assigned N and Observed N when you are SURE that there is no attrition.

F11. Assigned N for the comparison group (or pretest, if this is a pretest-posttest effect size) [CXNA]____
F12. Assigned N for the treatment group (or posttest, if this is a pretest-posttest effect size) [TXNA]____
F13. Observed N for the comparison group (or pretest, if this is a pretest-posttest effect size) [CXNO]____
F14. Observed N for the treatment group (or posttest, if this is a pretest-posttest effect size) [TXNO]____

Other Effect Size Data Fields

Enter these in the appropriate effect size data fields in CMA.
REFERENCES


Evaluating evidence-based programs and capacity-building initiatives to reduce school violence and drug use: Pushing the boundaries of prevention research. In M. Ortiz, & C. Rubio (Eds.), *Education Evaluation: 21st century issues and challenges* (pp. 367-405). New York.

To your health: How to understand what research tells us about risk. New York: Oxford University Press.


Implementing Second Step: Balancing fidelity and program adaptation. *Journal of Educational and Psychological Consultation*, 17, 1-29. doi: 10.1080/10474410709336588


The use of digital storytelling to improve the effectiveness of social and conflict resolution skill training for elementary students. *ProQuest LLC*.


Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Institute of Education Sciences.


Personal communication with Eddie Daggett, client support services from CFC, May/June 2013 (see gmoy.17@gmail.com).


*Asterisk indicates study included in meta-analysis.*
VITA

Gregory Moy was born and raised in Chicago, IL. Before attending Loyola University Chicago, he was employed at the University of Illinois-Chicago as Coordinator of Clinical Research and Programs in the Department of Psychiatry. He earned a Bachelor of Arts in Neuroscience and Psychology in 2003 from Macalester College.

While at Loyola, Moy served as National Association of School Psychologists’ Student Leader and became actively involved with several work groups and committees at state and national levels. Moy is currently completing a pre-doctoral internship with the Winnetka Public Schools. He has accepted a position as the school psychologist at The Joseph Sears School in Kenilworth, IL and continues to serve as adjunct faculty at Loyola University Chicago.