This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

Jump down to document ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

Browse Books & Publications
Make a charitable contribution

For More Information

Visit RAND at www.rand.org
Explore Pardee RAND Graduate School
View document details

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Unauthorized posting of RAND PDFs to a non-RAND Web site is prohibited. RAND PDFs are protected under copyright law. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please see RAND Permissions.
This product is part of the Pardee RAND Graduate School (PRGS) dissertation series. PRGS dissertations are produced by graduate fellows of the Pardee RAND Graduate School, the world’s leading producer of Ph.D.’s in policy analysis. The dissertation has been supervised, reviewed, and approved by the graduate fellow’s faculty committee.
No Bullies Allowed

Understanding Peer Victimization, the Impacts on Delinquency, and the Effectiveness of Prevention Programs

Jennifer S. Wong

This document was submitted as a dissertation in March 2009 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Peter W. Greenwood (Chair), Matthias Schonlau, M. Rebecca Kilburn, and Rosalie L. Pacula.
The Pardee RAND Graduate School dissertation series reproduces dissertations that have been approved by the student’s dissertation committee.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND’s publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.
ABSTRACT

Over the past decade school bullying has emerged as a prominent issue of concern for students, parents, educators, and researchers in North America and around the world. Research evidence suggests non-trivial and potentially serious negative repercussions of both bullying and victimization. The first chapter of the dissertation presents a comprehensive narrative literature review on the nature and significance of school bullying including controversies in definition; types of measurement; a description of victims, bully-victims, bullies, and bystanders with regard to defining characteristics, risk and protective factors, and outcomes and correlates of involvement in bullying/victimization; and a review of theoretical models that explain why bullying occurs. The second chapter of the dissertation uses a large, nationally representative panel dataset, the NLSY97, and a propensity score matching technique to assess the impact of bully victimization on a range of 10 delinquency outcomes measured over a six-year period. This analytic strategy considers the effect of baseline group differences by matching bullied and non-bullied subjects on propensity scores, thus allowing observable covariates to be eliminated as potential confounders of the estimated treatment effect. Results show that victimization prior to the age of 12 years is significantly predictive of the development of several delinquent behaviors, including running away from home, selling drugs, vandalism, theft, other property crimes, and assault. Using meta-analysis, the final chapter of the dissertation assesses the overall effectiveness of school-based programs for preventing bullying and victimization. Results suggest that as a whole, prevention programs are significantly effective at reducing the problem of victimization in schools, but are only marginally successful at reducing bullying. After participating in bullying prevention programs, students report an effect size of .188 for reduction in victimization, and an effect size of .109 for reduction in bullying others. The possibility of systematic between-study heterogeneity was explored via moderator analyses, and several significant moderators of treatment impact on victimization were identified. More work is needed to determine why programs are more successful with victims of bullying than with perpetrators, and prevention efforts should focus on the development of programs that are more likely to bring about successful reductions in both bullying and victimization.
# Table of Contents

Table of Contents v  
List of Tables xi  
List of Figures xiii  
 Acknowledgments xv  

Chapter 1 The Nature and Significance of School Bullying: A Narrative Synthesis 1  
  Introduction .......................................................................................................................................... 1  
  Research on School Bullying .................................................................................................................. 3  
  Implications of the Interest in School Bullying ................................................................................. 6  
  Methods for the Literature Review .................................................................................................... 6  
 Bullying Defined .................................................................................................................................. 7  
  Challenges in Adopting a Definitional Consensus ........................................................................... 8  
  The Tri-Criterion Definition ................................................................................................................ 10  
 Types of Bullying Behaviors ............................................................................................................... 12  
    Direct Bullying .................................................................................................................................. 13  
    Indirect Bullying ................................................................................................................................. 14  
    Behaviors Related to Bullying .......................................................................................................... 17  
 Prevalence of Bullying and Victimization ............................................................................................ 18  
    Prevalence in the United States ......................................................................................................... 20  
    Prevalence in Other Countries .......................................................................................................... 22  
      Relevance for American Populations ......................................................................................... 23  
 Measurement ...................................................................................................................................... 26  
    Time Frame of Assessment ................................................................................................................ 26  
    Types of Measures ............................................................................................................................. 27  
      a) Self-Report Questionnaires ................................................................................................. 27  
      b) Interviews and Focus Groups ................................................................................................. 30  
      c) Teacher Reports ......................................................................................................................... 31
d) Peer Nominations and Peer Relations Surveys ................................................................. 33

e) Direct Observations ........................................................................................................ 35

f) Biological Measures ........................................................................................................ 37

Concordance of Measurement Techniques ........................................................................ 37

Factors Affecting Measurement Correspondence ................................................................ 38

Summary and Conclusion on Measurement ......................................................................... 39

Four Participant Roles: Bullies, Victims, Bully-Victims, and Bystanders ............................... 42

I. Victims (Submissive) ........................................................................................................... 42

Characteristics of Submissive Victims .................................................................................. 42

Stability of Victim Status ...................................................................................................... 46

Victim Risk and Protective Factors ..................................................................................... 49

Reporting and Reactions of Others ..................................................................................... 53

Victim Reactions to Bullying ............................................................................................... 55

Outcomes and Correlates of Victimization .......................................................................... 58

II. Bully-Victims/Provocative Victims .................................................................................. 71

Characteristics of Bully-Victims .......................................................................................... 71

Prevalence of Bully-Victims ................................................................................................. 72

Bully-Victim Risk and Protective Factors .......................................................................... 73

Outcomes and Correlates of Bully-Victimization ................................................................ 74

III. Bullies (Perpetrators) ....................................................................................................... 77

Aggressive versus Passive Bullies ....................................................................................... 77

Characteristics of Bullies ....................................................................................................... 78

Persistence of Bullying Behaviors ......................................................................................... 85

Bully Risk and Protective Factors ....................................................................................... 85

Outcomes and Correlates of Bullying .................................................................................. 90

IV. Bystanders .......................................................................................................................... 94

Outcomes and Correlates for Bystanders ............................................................................ 95

Why Bullying Occurs – Theoretical Models ....................................................................... 96
Control/Matching Variables ................................................................. 134
Treatment of Missing Data ................................................................. 137
Results .................................................................................................. 139
Descriptive Results for Treatment Variable and Covariates............ 139
Unadjusted Results for Dependent Variables ................................. 142
Adjusted Results Using Logistic Regression .................................... 145
Propensity Score Estimation ............................................................... 148
Covariate Balance Before and After Propensity Score Matching .... 149
Results from Propensity Score Matching .......................................... 153
 Common Support .............................................................................. 153
 Treatment Effect Estimates ............................................................ 153
Comparison of Regression versus Propensity Score Matching Estimates .... 156
Assessing the Robustness of Findings .............................................. 157
Changes to Propensity Score Matching Model............................... 157
Changes to Dependent Variable Measurement .............................. 159
Discussion ......................................................................................... 161
Limitations and Future Research ..................................................... 164
Policy Implications .......................................................................... 166
Chapter 3 Are School-Based Bullying Prevention Programs Effective? A Meta-Analytic Review 168
Introduction ...................................................................................... 168
The Problem of School Bullying ...................................................... 168
Bullying Prevention Efforts ............................................................... 169
Evidence-Based Prevention ............................................................. 170
The Current Study ........................................................................... 170
Previous Reviews of Bullying Prevention Programs ....................... 171
Background on Meta-Analysis ......................................................... 175
What is Meta-Analysis? ................................................................. 175
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths of Meta-Analysis</td>
<td>175</td>
</tr>
<tr>
<td>Weaknesses of Meta-Analysis</td>
<td>177</td>
</tr>
<tr>
<td>Data Extraction</td>
<td>178</td>
</tr>
<tr>
<td>Comparability of Study Outcomes</td>
<td>178</td>
</tr>
<tr>
<td>Effect Size Calculations</td>
<td>182</td>
</tr>
<tr>
<td>Extracting Other Data</td>
<td>189</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>190</td>
</tr>
<tr>
<td>Pooled Analysis</td>
<td>190</td>
</tr>
<tr>
<td>The Homogeneity Statistic ($Q$)</td>
<td>192</td>
</tr>
<tr>
<td>Fixed Effects versus Random Effects Models</td>
<td>193</td>
</tr>
<tr>
<td>Modeling Between-Study Differences</td>
<td>195</td>
</tr>
<tr>
<td>Sensitivity Analysis</td>
<td>201</td>
</tr>
<tr>
<td>Methods</td>
<td>203</td>
</tr>
<tr>
<td>Objectives of the Current Study</td>
<td>203</td>
</tr>
<tr>
<td>Identification and Selection of Studies</td>
<td>205</td>
</tr>
<tr>
<td>Search Process</td>
<td>205</td>
</tr>
<tr>
<td>Selection Criteria</td>
<td>206</td>
</tr>
<tr>
<td>Independence of Effect Sizes</td>
<td>209</td>
</tr>
<tr>
<td>Effect Size Calculations</td>
<td>211</td>
</tr>
<tr>
<td>Estimation of Missing Data</td>
<td>211</td>
</tr>
<tr>
<td>Effect Size Formulas</td>
<td>212</td>
</tr>
<tr>
<td>Cluster Adjustments</td>
<td>215</td>
</tr>
<tr>
<td>Other Data Coded</td>
<td>215</td>
</tr>
<tr>
<td>Software</td>
<td>216</td>
</tr>
<tr>
<td>Results</td>
<td>216</td>
</tr>
<tr>
<td>Characteristics of the Included Studies</td>
<td>216</td>
</tr>
<tr>
<td>Meta-Analysis of Bullying Outcomes</td>
<td>220</td>
</tr>
<tr>
<td>Pooled Results</td>
<td>220</td>
</tr>
</tbody>
</table>
List of Tables

Table 1 Common Types of Bullying Behaviors................................. 13
Table 2 Characteristics of Victims as Rated by Experts................................. 46
Table 3 Characteristics of Bullies as Rated by Experts................................. 84
Table 4 Reasons for Bullying........................................................................ 97
Table 5 Screening and Target Sample Sizes................................................ 117
Table 6 NLSY97 Round 1 Sample Racial/Ethnic and Gender Composition...................... 117
Table 7 Response Rates by Racial/Ethnic and Gender Composition in Rounds 1–7........ 120
Table 8 Control/Matching Variables (1997 Baseline Survey)........................ 136
Table 9 Missing Data for Control/Matching Variables .................................... 138
Table 10 Victims of Bullying (1997 Baseline Survey)....................................... 139
Table 11 Baseline Demographics of Bullied and Non-Bullied Youths.................. 140
Table 12 Baseline Demographics of Respondents versus Non-Respondents........ 142
Table 13 Unadjusted Results for Dependent Variables across Total Sample........ 143
Table 14 Unadjusted Results for Dependent Variables by Treatment Group Status ...... 144
Table 15 Unadjusted and Adjusted Odds Ratios from Logistic Regressions of Bully
Victimization on Delinquency Outcomes....................................................... 147
Table 16 Logistic Regression of Treatment (Bully Victimization) on Covariates........ 148
Table 17 Balance of Covariates for Bullied and Non-Bullied Youths Pre- and Post-
Adjustment ..................................................................................................... 151
Table 18 Common Support for Treated Subjects............................................. 153
Table 19 Propensity Score Matching Estimates............................................. 154
Table 20 Magnitude of Effects for Significant Outcome Variables........................ 155
Table 21 Comparison of Regression and Propensity Score Matching Estimates........ 157
Table 22 Propensity Score Model Options for Outcome of Assault.................... 158
Table 23 Additional Model Options for Outcome of Assault............................ 159
Table 24 Propensity Score Matching Estimates Using Truncated Dependent Variables...... 160
Table 25 Treatment Effect Estimates for 6-Wave versus 3-Wave Dependent Variables ...... 161
Table 26 Conclusions for Logistic Regression and Propensity Score Matching Estimates .... 163
Table 27 Delinquent Behavior Cases Attributable to Bully Victimization ......................... 167
Table 28 Previous Narrative Reviews and Meta-Analyses of Bullying Prevention Programs . 174
Table 29 Characteristics of Studies Included in the Meta-Analyses .................................. 218
Table 30 Characteristics of the School-Based Bullying Prevention Programs ..................... 220
Table 31 Study-Level Data for the Meta-Analysis of Bullying Outcomes ............................ 223
Table 32 Fixed Effects Meta-Analysis for Bullying Outcomes ......................................... 224
Table 33 Subgroup Analysis of Study Characteristics on Bullying Outcomes ..................... 228
Table 34 Subgroup Analysis of Treatment Elements on Bullying Outcomes ....................... 229
Table 35 Tests for Publication Bias in Bullying Outcomes Data Set ................................. 231
Table 36 Influence of Individual Studies on Pooled Treatment Effect on Bullying
Outcomes ................................................................................................................................. 234
Table 37 Study-Level Data for the Meta-Analysis of Victimization Outcomes ..................... 237
Table 38 Fixed Effects Meta-Analysis for Victimization Outcomes .................................... 238
Table 39 Subgroup Analysis of Study Characteristics on Victimization Outcomes ............... 240
Table 40 Subgroup Analysis of Treatment Elements of Victimization Outcomes ................. 241
Table 41 Meta-Regression on Victimization Outcomes ...................................................... 243
Table 42 Tests for Publication Bias in Victimization Outcomes Data Set ............................. 244
Table 43 Influence of Individual Studies on Pooled Treatment Effect on Victimization
Outcomes ................................................................................................................................. 247
Table 44 Key Study Characteristics ....................................................................................... 343
Table 45 Characteristics of the Bullying Prevention Programs (Study-Level) ...................... 345
Table 46 Study-Level Data for Fixed Effects Meta-Analysis of Bullying Outcomes .......... 346
Table 47 Study-Level Data by Research Design Subgroup for Bullying Outcomes Meta-
Analysis ................................................................................................................................. 351
List of Figures

Figure 1 Selection of NLSY97 Respondents.............................................................................................................. 118
Figure 2 Unadjusted Results for Dependent Variables by Treatment Group Status ............................... 145
Figure 3 Box Plot of Propensity Scores for Bullied versus Non-Bullied Youthsa ............................. 149
Figure 4 Galbraith Plot of Study Heterogeneity for Bullying Outcomes ......................................................... 222
Figure 5 Forest Plot of Study Effect Sizes for Bullying Outcomesa ................................................................. 224
Figure 6 Begg's Funnel Plot of Study Effect Size vs. Standard Error, Bullying Outcomes ......... 231
Figure 7 Egger's Publication Bias Plot of Standardized Effect Size Estimates vs. Precision, Bullying Outcomes ........................................................................................................................................... 232
Figure 8 Individual Study Influences on Pooled Estimate of Treatment on Bullying Outcomes .......................................................................................................................................................... 235
Figure 9 Galbraith Plot of Study Heterogeneity for Victimization Meta-Analysis ............................. 236
Figure 10 Forest Plot of Study Effect Sizes for Victimization Outcomesa ............................................... 239
Figure 11 Begg's Funnel Plot of Study Effect Size vs. Standard Error, Victimization Outcomes ............................................................................................................................................. 245
Figure 12 Egger's Publication Bias Plot of Standardized Effect Size Estimates vs. Precision, Victimization Outcomes ............................................................................................................................................. 246
Figure 13 Individual Study Influences on Pooled Estimate of Treatment on Victimization Outcomes ............................................................................................................................................. 248
Acknowledgments

I admit – this acknowledgements page is the most enjoyable thing I have ever written! The dissertation process was far longer and more challenging than I ever expected, and at many points felt interminable. To have reached the end of this journey feels surreal.

I did not cross the finish line without the assistance of several individuals who offered me tireless support, encouragement, and gallons of coffee. First and foremost, gratitude to my dissertation committee: Peter Greenwood, Matt Schonlau, Becky Kilburn, and Rosalie Pacula. Matt, I am forever grateful to you for joining my committee when I was desperate, and for your perfect blend of excellent advising while leaving me in complete ownership of my analysis. Your commitment to helping me finish this project was far, far beyond expectations. Becky, many thanks for helping me to conceptualize the original plan, for the incredibly generous subcontract through PPN, and for your eyes that see things mine do not. Thank you to Rosalie, for always being responsive and for offering invaluable advice and feedback despite never having worked with me before! And thanks to Peter, for your faith in my abilities and for being the original inspiration who forever changed the course of my academic work by introducing me to the world of Criminology.

I deeply appreciate the guidance from my external reader Dr. Mark Lipsey, whose work I have admired for years, and who graciously provided me with detailed and enormously helpful comments (and answers to my excessive follow-up questions). Also, thank you to the National Center for School Engagement at The Partnership for Families & Children, for generously supporting a portion of this dissertation.

A mountain of thanks to John, for sticking by me through the circus, for believing I'm "a problem but worth it", and for the hours upon hours of photocopying in the UCLA and USC stacks (properly outfitted with sunglasses and iPod). Thank you to Mum, Big, Danny, and Andrew – for putting up with your stressed, impoverished, grouchy, MIA daughter/sister for far too long. Thanks are also due to Beverly Maxwell, for understanding the unique pain of being ABD, and for continually reminding me how good it is on the other side. Thank you to
Heather Peacock Bawa, for always being a supportive friend no matter how often I failed to return phone calls and emails (and for continuing to leave me messages!) Also, thanks to my friends and colleagues at Simon Fraser University, for your support, advice, and patience with my never-ending struggle, and for having the discretion not to ask the dreaded "so how much longer until you're done?!" And last but not least, thank you to my friend and original mentor Laura Hickman, for your unlimited wisdom, for always being a problem solver, and for teaching me how to "be" a researcher.
Chapter 1
The Nature and Significance of School Bullying: A Narrative Synthesis

INTRODUCTION

The interactions that children have with their peers are critical for helping them to develop social skills, acquire norms for appropriate and inappropriate interpersonal conduct, and derive standards against which to measure and compare their own behavior. Schools are environments where students congregate to gain knowledge, practical skills, and social skills aimed at improving success in post-secondary life. Yet the acquisition of social skills is not likely to be simple and painless. Not all peer interactions are positive or instructive, and both overt and covert forms of aggression are commonplace in the school setting. Some research suggests that aggressive behavior is learned early in life and becomes resistant to change if it continues past the child's eighth year (e.g., Huesmann, Eron, Lefkowitz, & Walder, 1984).

School bullying, a subtype of aggression (Griffin & Gross, 2004; Pellegrini, 2002) that may or may not include interpersonal violence, has emerged in recent years as a topic of focus in the fields of education, criminal justice, and public health. The problem is widespread, both in the United States and around the world, as evidenced by a sizeable collection of international research – most notably from Australia, Canada, the United Kingdom, Japan, and Scandinavia (e.g., Craig, Pepler, & Atlas, 2000; Rios-Ellis, Bellamy, & Shoji, 2000; Salmivalli, Lagerspetz, Bjorkqvist, Osterman, & Kaukiainen, 1996b; Sharp & Smith, 1994; Slee, 1995). In the United States a nationally representative sample of over 15,600 students in grades six through ten found that nearly 30 percent of students reported moderate or frequent involvement in bullying at school, with 13 percent participating as a bully, 11 percent as a victim, and 6 percent as both bully and victim (Nansel et al., 2001). These findings suggest that over 16 million students in the U.S. are involved in bullying to some capacity.1

---

1 Based on 2007 school enrollment data provided by the United States Census Bureau (2007).
Bullying among children is not a new phenomenon, and forms of bullying have likely existed for as long as there have been schools. Indeed, bullying is often considered a normal developmental experience, or rite of passage, for children in school (Smith & Brain, 2000). Almost all adults have distinct childhood memories of occasions when they bullied another child, were victimized themselves, or witnessed a bullying situation. Many adults, regardless of age, are able to discuss details of the events, such as the name of the bully, their grade level at the time of the incident, and the location in which the bullying occurred (Barone, 1997). For example, Ambert (1994) examined 180 Canadian college students' autobiographies for their depictions of happy and unhappy childhood events. Twenty-seven percent of the students in one cohort and 17 percent of the students in another cohort relayed stories of peer abuse described as having had a lasting detrimental effect. The students' descriptions of peer abuse were long and detailed, and many reported physical illness, mood disorders, and problems with schoolwork as a result of their experiences. In another study, Crozier and Skliopidou (2002) studied a sample of 220 adults with respect to their recollections of name-calling while at school. Two-thirds of the adults recalled being teased with hurtful names, and reported feelings of anger, unhappiness, and shame as a result. Twenty-four percent of the sample indicated that the memories of name-calling were still very painful, and believed the experiences had produced long term negative effects on their personality and attitudes.

Unfortunately, myths abound that bullying is a "normal" part of growing up, that learning to stand up to a bully is productive because it "builds character," and that almost everyone experiences bullying without long-term negative effects. As discussed by Rigby (1999b):

*For many people bullying has not figured in their lives to any troublesome degree. Many men and women who were bullied at school have largely forgotten about it and have become preoccupied with something else. Occasionally, one meets somebody who is inclined to brag about being bullied at school and how through some supreme act of heroism he (usually he) managed to turn the tables on the bastard. One gentleman I know wrote to a West Australian newspaper to say that he had been bullied at school as a skinny little kid, but had responded by building up his strength and*
determination and, behold he was now heavy weight boxing champion of Canada. Bullying, he proclaimed, can do you good. Moreover, there is at large a good deal of scarcely veiled contempt for the wimpish kid who allows himself or herself to be bullied. Hence for one reason or another, there remains a certain resistance to the proposal that bullying should be stopped on the grounds that it does significant harm to people. Scepticism about the negative effects of bullying may well extend to those who have the power to act so as to counter bullying in schools and beyond. (p. 2)

Despite these lingering viewpoints, research evidence on the negative and potentially serious outcomes for victims of bullying and bullies themselves has helped sway apathetic attitudes towards this pervasive problem. Even if the existence of bullying is typical for schools, condoning the behavior or accepting it as unavoidable is no longer the corollary.

Research on School Bullying

While the earliest study on school bullying was published by Charles Vaughn in 1941, wherein the author reports that students with weak reading skills are likely to bully other students, the vast majority of research on bullying has been conducted in the past decade. Bullying first piqued the interest of American researchers, the media, and the general public in the early 1990s, with a surge in the amount of published research beginning around 1999.2 The United States has not been a frontrunner in bullying research, entering into the field behind the Scandinavian countries, England, Australia, and Japan. The first efforts to rigorously study school bullying were conducted in the 1970s by Dan Olweus in Sweden and Norway, resulting in the publication of the well-cited book, Aggression in the Schools: Bullies and Whipping Boys (Olweus, 1978). The Scandinavian public became engrossed with the bullying phenomenon in the early 1980s, when three young Norwegian boys committed suicide after suffering persistent bullying by peers (Olweus, 1993c). A national campaign against bullying ensued, which led to the creation of the Olweus Bullying Prevention Program – now an

---

2 Key bullying researchers in the United States include Nicki Crick, Dorothy Espelage, Sandra Graham, Jaana Juvonen, Becky Kochenderfer-Ladd, Anthony Pellegrini, David Schwartz, Eric Storch, and Stuart Twemlow.
international and widely implemented model for schools (Olweus, Limber, & Mihalic, 1999). Olweus, widely considered the father of bullying research, has published extensively on the subject for almost 40 years.

Following the original Norwegian work, a 1987 conference on bullying was hosted in Stavanger, Norway, followed by programs of research in countries such as Finland (see research by authors Riittakerttu Kaltiala-Heino, Kirsti Kumpulainen, and Christina Salmivalli) and the United Kingdom (key researchers include Michael Boulton, Helen Cowie, Sonia Sharp, and Peter K. Smith). During the 1980s research on bullying also began in Japan, and increased following a rash of bullying-related suicides among students in the early 1990s (e.g., see research by Yoji Morita and Mitsuru Taki). Since the mid-1990s, substantial work on bullying has been conducted in Australia (most notably by researchers Ken Rigby and Phillip Slee), in Canada (Wendy Craig and Debra Pepler), and in Italy (see work by Anna Baldry and Ersilia Menesini).

In the United States, research on bullying escalated dramatically following the deaths of 12 students and a teacher in the April, 1999 shooting rampage at Columbine High School, Colorado. The incident at Columbine was the most devastating in terms of the number of deaths, but was by no means the first episode of school shooting – or the last. At least seven incidents took place in the preceding school year, while eight similar though less deadly incidents followed over the course of the next two years (InfoPlease Almanac, 2006). Columbine garnered an enormous amount of media coverage, and captivated public attention for several weeks. Early anecdotal reports surfaced claiming that shooters Dylan Klebold and Eric Harris had been frequent targets of bullying by their peers (Limber, 2002). A subsequent investigation of 37 school shooting incidents across 26 states reported that over two-thirds of the shootings were carried out by attackers who had been “bullied, threatened, attacked, or injured by others prior to the incident ... in a number of cases, attackers described experiences of being bullied in terms that

3 In the two years following the Columbine shootings, incidents took place in Conyers, Georgia (May, 1999; 6 wounded); Deming, New Mexico (November, 1999; 1 killed); Fort Gibson, Oklahoma (December, 1999; 4 wounded); Mount Morris Township, Michigan (February, 2000; 1 killed); Savannah, Georgia (March, 2000; 2 killed), Santee, California (March, 2001; 2 killed); Williamsport, Pennsylvania (March, 2001; 1 wounded); and Granite Hills, California (March 2001; 4 wounded).
Similarly, in a study of all school-associated violent deaths in the United States between July 1, 1994 and June 30, 1999, homicide perpetrators were found more than twice as likely as homicide victims to have been bullied by peers (19.8 percent versus 8.8 percent; Anderson et al., 2001).

While these studies do not present causal assessments of the impact of bully victimization on school homicides, they strongly suggest that victimization by peers is linked to dangerous outcomes.

In addition to the potential link with extreme violence and homicide, a number of studies suggest that victimization by bullies, as well as perpetration of bullying, may be correlated with adolescent suicide (e.g., Kim, Koh, & Leventhal, 2005; Olweus, 1993a; Smith, Talamelli, Cowie, Naylor, & Chauhan, 2004b). For example, Kaltiala-Heino, Rimpela, Marttunen, Rimpela, and Rantanen (1999) studied 16,510 Finnish youths aged 14 to 16 years and found that both bullies and victims were more likely to be depressed and to have suicidal ideations than were non-involved youths. Suicide is the third leading cause of death among adolescents in the United States (Centers for Disease Control and Prevention, 2004), and research suggests that rates of attempted suicide are much higher. For example, a study of 15- to 19-year-old adolescents using a nationally representative sample of U.S. emergency departments found that for every suicide death, 32 youths were treated for self-inflicted injuries (Ikeda et al., 2002).

Non-fatal outcomes for victims and perpetrators of bullying are far more commonplace. In a study by Hoover, Oliver, and Hazler (1992), over 14 percent of middle- and high-school students reported being severely traumatized or distressed by bullying, while 16 percent suffered negative impacts with respect to their academic performance. Research by Garrity, Jens, Porter, Sager, and Short-Camilli (1997) finds that 20 percent of students are scared throughout much of the school day, while a report by the National Center for Education Statistics indicates that five percent of students miss school or avoid particular locations on school grounds because they are afraid (DeVoe, Peter, Noonan, Snyder, & Baum, 2005). In addition, a study by the Kaiser Family Foundation (2001) finds that more 8- to 15-
year-olds select teasing and bullying as "big problems" in their lives than issues such as drugs and alcohol, violence, racism, AIDS, or pressure to have sex.

Implications of the Interest in School Bullying

As a result of research findings, media attention, and public concern, bullying has become a salient political issue. Anti-bullying legislation has been passed in 35 U.S. states that in some cases may hold individuals and schools accountable for costs and damages that students suffer as the result of school bullying (Srabstein, Berkman, & Pyntikova, 2008). Recent court rulings highlight the potential implications for schools and perpetrators of bullying behaviors. For example, after a bullied boy in Yokohama, Japan committed suicide, his parents won record-breaking damages from the local education authority (39.5 million yen, or approximately $387,000 U.S. dollars; Fitzpatrick, 2001). Nine former students of the victim's junior high school were also ordered to pay damages of approximately $18,400 U.S. dollars. In a case involving a child who was punched and intimidated by a gang and subsequently refused to go to school, a suit was filed against a San Francisco school system, charging that the school district, teacher, and principal failed to provide the child’s right to a safe, secure, and peaceful school environment (Ballard, Argus, & Remley Jr., 1999). Meanwhile in British Columbia, Canada, a teenage girl was convicted of criminal harassment and uttering threats, which were judged to have contributed to the suicide of a 14-year-old girl (Greenfield, 2002a). In a separate case, a human rights tribunal in British Columbia awarded a student plaintiff a financial settlement after ruling that the school board discriminated against him by failing to provide protection from homophobic bullying (Greenfield et al., 2002b).

Methods for the Literature Review

The current chapter consists of a comprehensive literature review on the topic of school bullying, inclusive to September 18, 2006, restricted to studies published in the English language but without limits on publication date. Fourteen databases were included in the search: Criminal Justice Abstracts, Dissertation Abstracts, Econlit, Education Abstracts,
BULLYING DEFINED

What does it mean to "bully" or "be bullied"? Understanding and measuring bullying is more complex than it may first appear. One of the difficulties in measurement stems from the multitude of definitions that exist and because researchers, policymakers, students, teachers, parents, media, and the general public assume they already know what bullying is. A considerable challenge in bullying research is the loose definition of the term, which frequently differs among studies and without question has changed in recent years. In addition to "bullying," the phenomenon is often termed "peer victimization" (e.g., Juvonen & Graham, 2001; Kochenderfer & Ladd, 1996a), "peer harassment" (e.g., Juvonen, Nishina, & Graham, 2000; Newman, 2003), and "peer abuse" (e.g., Olweus, 1993b). A common assumption is that bullying denotes behavior that is physical and overt (e.g., hitting, kicking, or pushing another child), with the stereotypical example depicting a large boy shoving a smaller boy into a locker or stealing his lunch money. While such behaviors are indeed forms of bullying, other forms include direct but non-physical behaviors such as taunting and intimidating, as well as indirect methods such as ostracizing, posting anonymous insults on a web page, and spreading rumors. Some studies suggest that verbal bullying is the most common type of bullying experienced by children and youths (e.g., Bentley & Li, 1995; Boulton, Trueman, & Flemington, 2002; Rivers & Smith, 1994).
Challenges in Adopting a Definitional Consensus

Bullying is not akin to a clear-cut behavior such as carrying a weapon in school, and the lack of definitional consensus makes it difficult to develop policy. Limber and Small (2003) compared the definitions of bullying across 19 states and found substantial differences. For example, Colorado defines bullying as "any written or verbal expression, or physical act or gesture... intended to cause distress upon one or more students" (C.R.S., 2001). New Jersey defines bullying as a behavior that "disrupts both a student's ability to learn and a school's ability to educate its students in a safe environment" (N.J.S.A., 2002), while Georgia defines bullying as "any willful attempt or threat to inflict injury on another person ... or any intentional display of force such as would give the victim reason to fear or expect immediate bodily harm" (Ga. Code Ann., 2001).

Srabstein, Berkman, and Pyntikova (2008) reported that as of June, 2007, 35 U.S. states had passed anti-bullying laws. Twenty-five of these statutes provided a definition of bullying, but definitions ranged substantially across states. As discussed by Srabstein et al., A number of statutes characterized bullying as a verbal, written, physical act or gesture, caused by one or more students against another student. Some laws require that the acts of bullying be "overt," "repeated," or "severe, persistent, and pervasive." Other laws define bullying in terms of mental state, requiring an understanding of its purpose, motivation, or cause, or deleterious effects (such as inflicting physical or emotional harm, or creating a hostile and intimidating environment). Finally, some statutes refer to different locations where bullying can occur, such as school grounds and vehicles; various situations, such as school activities or sanctioned events; or through the use of telephones or computers. (p. 13)

When definitions of bullying are compared across empirical studies, it is clear that researchers have not yet reached a consensus regarding what exactly bullying entails and how it should be defined and measured. This is especially true with older studies and studies conducted outside of the United States, and it presents a problem when attempting to compare frequency or prevalence rates across multiple reports. Elinoff, Chaffouleas, and Sassu (2004) lament that the bulk of the literature on bullying uses the term "bullying" as if it were
operationally defined. Even the list of 15 criteria used to make a clinical diagnosis of conduct disorder in the *Diagnostic and Statistical Manual of Mental Disorders* includes "bullies, threaten, or intimidates others" (American Psychiatric Association, 2000). Many studies on bullying do not clearly state the definition of bullying that was used during data collection, and misperceptions about the definition may lead to erroneous reports by respondents and subsequent misinterpretations of research findings. Researchers admit the challenges of developing a definition which adequately captures all of the elements of bullying and incorporates the many ways in which this aggression may take form (Nansel & Overpeck, 2003).

As such, the varying definitions of bullying used by researchers may lead to biased estimates of child involvement – although the direction of bias is not always clear. For example, Griffin and Gross (2004) observe that while some studies consider any and all intentionally aggressive behavior toward others to be bullying, other studies are much more specific in defining the types of behaviors that constitute bullying and subsequently report much lower prevalence rates. Evidence suggests that respondents have firm, pre-conceived beliefs regarding the definition of bullying, which may or may not be accurate. For example, when 15-year-old boys were asked how much bullying occurred in their schools, the majority of respondents initially indicated that there was none (Thompson & Arora, 1991). When the subjects were subsequently asked whether there was any verbal bullying at their schools, they all responded affirmatively and were able to give examples. This discrepancy is particularly salient because immediately prior to the interviews the subjects were provided with a definition of bullying that included examples of both verbal and non-verbal bullying behaviors. This finding suggests that the subjects' ex ante definitions of "bullying" held constant despite their exposure to a comprehensive definition.

---

4 Since three or more criteria are necessary for a diagnosis of conduct disorder, bullying alone is not sufficient for a diagnosis of this type.
The Tri-Criterion Definition

A frequently used definition of bullying contends that: "A student is being bullied or victimized when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other students" (Olweus, 1994, p. 1173). A definition by Hazler, Hoover, and Oliver (1992) notes that "bullying is a form of aggression in which one student – or a group of students – physically or psychologically harasses a victim over a long period of time" (p. 20), while Rigby (2002b) proposes that "bullying involves a desire to hurt + hurtful action + a power imbalance + (typically) repetition + an unjust use of power + evident enjoyment by the aggressor and generally a sense of being oppressed on the part of the victim" (p. 51). In the frequently used Olweus Bully/Victim Questionnaire, Dan Olweus provides a more lengthy definition of bullying, as follows:

We say a student is being bullied when another student, or several other students
- say mean and hurtful things or make fun of him or her or call him or her mean and hurtful names
- completely ignore or exclude him or her from their group of friends or leave him or her out of things on purpose
- hit, kick, push, shove around, or lock him or her inside a room
- tell lies or spread false rumors about him or her or send mean notes and try to make other students dislike him or her
- and other hurtful things like that.
When we talk about bullying, these things happen repeatedly, and it is difficult for the student being bullied to defend himself or herself. We also call it bullying, when a student is teased repeatedly in a mean and hurtful way. But we don’t call it bullying when the teasing is done in a friendly and playful way. Also, it is not bullying when two students of about the same strength or power argue or fight. (Olweus, 1996)

In general, the comprehensive definitions agree that to be defined as true bullying behavior, the aggressive acts by an individual or group of individuals must meet three criteria (e.g., Hoover & Oliver, 1996; Olweus, 1993c; Rigby, 2004). The behavior must:

a. be intended to cause harm or distress,
b. occur repeatedly over time, and
c. occur in a relationship in which there is an imbalance of power or strength.

Bullying behavior is differentiated from other types of aggressive behavior and conflict mainly as it pertains to the misuse of power (Betlem, 2001). The bully must be more powerful than the victim – physically, verbally, or socially – and the intent to hurt must be conscious; accidental physical or emotional bruising is not bullying. The repeated incidents of harassment are important in establishing a pattern of behavior as opposed to an isolated incident of aggression, and serve to reinforce the imbalance of power between the bully and victim. Elinoff and colleagues (2004) suggest that the definition should also include a requirement that the bully's hostile behavior is undertaken in the absence of provocation, rather than in response to actions by others. It bears repeating that it is not considered bullying when children of equal power and ability to defend themselves argue or fight; such fist-fighting between two boys of equal size, or gossip and rumor-spreading between two girls of similar social status. Furthermore, play-fighting or playful teasing among friends is not considered bullying if the intent is not malicious. The line between teasing and bullying is somewhat subjective and may be difficult to judge (Farrington, 1993; Hazler, Miller, & Carney, 2001).

Consensus on the tri-criterion definition is not universal; Greene (2006) argues that the elements proposed by Olweus are not straightforward to define:

For example, psychological intent to hurt is an elusive concept, particularly if the bully justifies his or her actions based on an anti-victim attitude, thus minimizing or suppressing thoughts and feelings about the harm incurred (Perry, Williard, & Perry, 1990). What if the victim states, perhaps out of pride or in deference to peer group norms, that the actions of the bully really did not hurt (Phillips, 2003)?... Similarly, the determination of a power imbalance is not an easy one to make (Woods & Wolke, 2003). This is particularly true in the case of non-face-to-face or indirect bullying (Rigby, 2002b). For example, students who spread nasty rumors via the Internet do not necessarily have more power than the targets of such rumors. (p. 24)
Nonetheless, the tri-criterion definition of bullying appears to be the most comprehensive and well-accepted definition that exists, and as such is the recommended definition for use in future research on school bullying. The issue of a clear and complete definition is more than just pedantic; anti-bullying policies and programs can only be effective if they are developed to target the range of bullying behaviors that are known to occur. Furthermore, assessments of changes in prevalence rates over time and the success of anti-bullying efforts are futile if measurements differ substantially across the studies being compared.

**Types of Bullying Behaviors**

Bullying behavior can assume a wide range of actions; see Table 1 for a list of common forms of bullying. Unfortunately, the advent of new forms of communication technology has added to the list of ways in which children and youths can be victimized by others. Types of bullying are generally classified into two categories: direct and indirect, while behaviors are classified as physical, verbal, and non-verbal/non-physical. According to Olweus (1993c), direct bullying involves overt, relatively open attacks between a bully and victim, while indirect bullying is more subtle and can be executed by the bully without ever confronting their target.
### Table 1 Common Types of Bullying Behaviors*

<table>
<thead>
<tr>
<th></th>
<th>Direct bullying</th>
<th>Indirect bullying</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical bullying</strong></td>
<td>Hitting, kicking, shoving, poking, pinching, biting, tripping, spitting, hair pulling, locking someone in a room or locker, destruction or theft of property, sexual harassment or assault, ganging up or cornering, burning, poisoning, stabbing, shooting</td>
<td>Enlisting a friend or older sibling to assault someone on your behalf</td>
</tr>
<tr>
<td><strong>Verbal bullying</strong></td>
<td>Taunting, teasing, insulting, name-calling, threatening, intimidating, humiliating, ridiculing, sexual propositioning, demanding servitude/money/property</td>
<td>Spreading rumors, gossiping, talking about someone just loud enough to be overheard, using code names when talking about others</td>
</tr>
<tr>
<td><strong>Non-verbal/Non-physical bullying</strong></td>
<td>Making faces, making threatening or obscene gestures, exhibitionism, voyeurism, stalking</td>
<td>Excluding others from a group, ostracizing, manipulating or sabotaging friendships, sending mean or threatening e-mails/notes/text messages/instant messages, blackmailing, creating ratings of individuals' personal characteristics (such as race, ethnicity, religion, national origin, or sexual orientation), playing mean or embarrassing jokes, making prank telephone calls, writing mean things on desks, moving personal property</td>
</tr>
</tbody>
</table>

*Adapted from Rigby (1996)

**Direct Bullying**

Direct bullying behaviors are those most commonly understood to constitute bullying. Direct physical bullying involves actions such as hitting, kicking, shoving, and spitting, while direct verbal bullying includes behaviors such as taunting, name-calling, and intimidating. In addition, direct non-verbal/non-physical bullying may include actions such as threatening gestures, exhibitionism, and stalking.
Indirect Bullying

Indirect bullying, also known as relational aggression or social aggression, can be just as harmful to victims as the experience of direct bullying (Crick & Bigbee, 1998; Storch, Brassard, & Masia-Warner, 2003), and has been linked to a variety of negative psychological, social, and physical effects (Clarke & Kiselica, 1997; Farrington, 1993). Indirect bullying includes behaviors such as manipulating, spreading rumors, ostracizing, enlisting friends to assault or intimidate a victim, and sending threatening e-mails. Relational aggression is more difficult to detect and more challenging to control, and many school harassment policies focus on physical or direct forms of bullying that are reasonably obvious and straightforward to observe. Crick, Casas, and Mosher (1997) found evidence of relational aggression in students as young as 3- to 5-years-old, and report that even at the preschool age victims of relational aggression may experience negative social-psychological impacts. Some research suggests that teachers deem relational bullying to be far less serious than other forms of bullying. As a result, some teachers may have less empathy for victims and may be less likely to intervene and punish bullies for their actions (e.g., Bauman & Del Rio, 2006; Craig, Henderson, & Murphy, 2000; Hazler et al., 2001; Yoon & Kerber, 2003). Much research indicates that males tend to choose physical and direct bullying methods, while females are more prone to verbal and indirect bullying behaviors (e.g., Crick & Bigbee, 1998; Lagerspetz, Bjorkqvist & Peltonen, 1988; Olweus, 1997). While males may be naturally more physically aggressive than females, it may also be that boys and girls choose to engage in the form of bullying that they consider most injurious. A study by Galen and Underwood (1997) found that boys rated physical aggression as more hurtful than social aggression, while girls reported a greater level of harm from social aggression.

Several innovative means for indirect bullying have evolved in recent years, resulting in new fields of research, new problems for schools and parents, and countless numbers of bullies and victims. For example, teenagers in the 1980s and 1990s made creative use of the advent of three-way calling telephone capabilities to bully their peers. With this indirect bullying technique, an unsuspecting victim was manipulated into saying something derogatory
about a third party who was covertly listening on the line. This supposed "backstabbing" of the
third party would subsequently be used by the bullies as a reason to ostracize the victim or to
engage in direct retaliation. Over the past decade, internet and telecommunication technologies
such as e-mail, instant messaging, text messaging, and online networking (e.g., Facebook,
MySpace) have become extremely popular and are widely used among America's youth. In
2002, 99 percent of public schools in the United States had computers connected to the
Internet (Parsad, Jones, & Greene, 2005), and 97 percent of children aged 12 to 18 reported
using the Internet (UCLA Center for Communication Policy, 2003). Furthermore, a report
conducted by the Pew Internet & American Life Project found that 74 percent of the 17
million teenagers who surfed the Internet in 2000 used America Online's Instant Messenger
service (Lenhart, Rainie, & Lewis, 2001).

According to Franek (2005/2006), a "cyberbully" is an individual who purposely
misuses online technology to harass, intimidate, bully, or terrorize. Research on cyberbullying
has only begun to emerge, but it appears to be a growing problem in middle- and upper-
middle-class communities, especially in the middle school grades (Blair, 2003). Cyberbullying
can involve inappropriate, cruel, hateful, harassing or threatening e-mails (delivered to the
victim alone or simultaneously forwarded to other individuals), text messages (including being
bombarded with multiple messages from known or anonymous senders), instant messages,
web site posts, blogs, or even entire websites. Websites such as www.ihatetomcruise.com abound,
and youths with an ounce of internet savvy have been known to create similar websites for
disliked peers, on which they post humiliating stories, jokes, and spiteful comments. For
example, bullies in Fayetteville, Arkansas set up a page on the social networking site
Facebook.com titled Every One That Hates Billy Wolfe, using the network to ridicule and
victimize an unpopular student (Davis, 2008). On-line voting booths are another cybertool
that can be used to bully and harass victims. Benfer (2001) describes an on-line web voting
booth that was created by a group of New York students to determine which female student in
the Manhattan interschool system had the reputation of being the most sexually promiscuous.
The voting booth was named the Interschool Ho and was posted on Freevote.com, accumulating
a ranked list of 150 students. The site was shut down only after the Brooklyn District Attorney became involved with the case. Ybarra and Mitchell (2004) find that nearly 50 percent of youths who are targets of Internet cyberbullying are also the targets of conventional bullying, while over 50 percent of online bullies also engage in offline bullying. This finding implies that around half of the youths involved in cyberbullying as aggressors or victims (or both) are exclusively involved in cyberbullying, suggesting that the Internet presents a unique environment that greatly expands the pool of aggressors and victims. Beckerman and Nocero (2003) suggest that anonymity is a key factor in why cyberbullying is a growing form of harassment – sitting behind a computer gives students a sense of power and control that they do not feel when face-to-face with their peers. In such cases, the bullies become bolder and are likely to make more aggressive and daring comments than they would be willing to make in person.

A nationwide survey of 1,501 youth aged 10 to 17 years found that 19 percent of regular internet users had experienced bullying online (Ybarra & Mitchell, 2004). As commented by Willard (2006), "There is no escape for those being bullied in cyberspace: The victimization is ongoing" (p. 43). Cyberbullying is particularly disturbing because materials can be targeted and distributed not just to an individual victim, but schoolwide or even worldwide – once information is posted on the internet it is often impossible to retrieve or destroy. Part of the difficulty in controlling these types of bullying behaviors is that the taunts or threats are often sent from an anonymous source. Cyberbullies have also been known to impersonate victims on e-mail and the Internet, for example, by breaking into a victim's e-mail account and writing threatening or vicious messages to third parties (Blair, 2003; Franek, 2006; Willard, 2006).

Cellular phones with photo, video, and web accessibility are also used for bullying purposes. Examples include taking embarrassing pictures (e.g., in a bathroom or locker room) and e-mailing or threatening to circulate the photos to others. Another example, termed "happy slapping," involves a victim being assaulted while one of the bully's friends digitally records the event on a cellular phone. The video is then circulated among the peer group to further the victim's humiliation (Brough & Sills, 2005). In addition, videos can easily be posted
on the internet and then viewed by millions. In April, 2008, eight Florida teenagers were arrested after a video of them punching, kneeing, and slapping a 16-year-old girl surfaced. The teens were allegedly retaliating for insults the victim had posted on the Internet, and were planning to post video of the assault on the website YouTube. The eight teens, six of whom are girls, will be tried as adults for crimes including kidnapping, battery, and tampering with a witness (CNN, 2008).

Behaviors Related to Bullying

Some researchers consider sexual harassment a form of bullying (e.g., Coy, 2001; Stein, 1995, 1999); however this type of harassment is not frequently discussed in studies of bullying. This may be because sexual harassment involves different motivations on the part of the bully, and bully and victim profiles that are not necessarily analogous to those found in more typical forms of bullying. For example, Pellegrini (2002) suggests that the influence of bullying on sexual harassment is moderated by the level of interest that bullies have in heterosexual relationships, in other words, sexual harassers are likely to be males who are both bullies and who have a high interest in relationships with females. Many studies discuss bullying and sexual harassment as behaviors that are related but separate (e.g., Cavendish & Salomone, 2001; DeSouza & Ribeiro, 2005; Land, 2003). Some behaviors do overlap, such as spreading rumors about sexual behavior, making sexual gestures, and using derogatory names in regard to an individual’s supposed sexual orientation. In contrast with other types of bullying, sexual harassment is prohibited under the federal civil rights in education law, Title IX; numerous successful lawsuits have been pursued against schools that failed to prevent sexual harassment in the educational environment (Stein, 1995). Sexual harassment in schools is widespread, and is deemed a serious problem by students. A nationally representative survey of 2,064 public school students in grades 8 through 11 found that 81 percent of students had experienced some form of sexual harassment, and nearly half (47 percent) of the students who had experienced such harassment said the incident made them feel "very" or "somewhat" upset (American Association of University Women, 2001).
Stalking is another related behavior that is sometimes but not usually included in studies on bullying. McCann (2001) points to significant parallels between bullying and stalking, suggesting that stalking and bullying are overlapping forms of harassment because they are repetitive and obsessive in nature and are directed at an individual. Both bullying and stalking involve a need to control others, and frequently include threats and intimidation. According to McCann, "bullying might constitute a childhood subtype or precursor to stalking related behavior later in life" (p. 129).

PREVALENCE OF BULLYING AND VICTIMIZATION

Determining prevalence rates of bullying and victimization is essential for making comparisons across populations from different school districts, states, countries, and cultures, and for comparing among groups with different defining characteristics such as gender, socioeconomic status, and age. Prevalence rates also allow for comparisons over time to assess rates of change in behaviors, as well as to evaluate the effects of policies or prevention programs. Up until 2001, U.S. prevalence rates for bullying were approximations extrapolated from small-scale studies, with the majority of the larger-scale studies on bullying being conducted in Europe, Japan, and Australia. This fact is somewhat paradoxical given the extent of youth violence in America compared with other First World countries – e.g., the annual prevalence rate in 1992/1993 for serious violence among 16- to 17-year-olds was 32 percent for the U.S. compared with 18 percent for the U.K. (Satcher, 2001). Since the mid-1990s there have been scores of small-scale studies in the United States focusing on one to several schools, along with several studies conducted at the district or state level. While these studies (depending on the particular methodology used) may be generalizable to the specific communities and regions that were sampled, they are not particularly helpful for generating an overall prevalence rate for the country. Furthermore, these studies frequently employ different operational definitions of bullying (e.g., in which victimization includes/omits both direct and indirect bullying methods), focus on different age groups (e.g., primary school versus high school), and involve different data collection methods (e.g., self-report versus peer
nomination) with different time frames (e.g., past week versus past year). As should be expected, these differences in research design have sizeable impacts on study results. The small-scale studies report a large variance in prevalence estimates, ranging from bullying rates of 2.4 percent (Hunt, 2002) to 19.8 percent (Berthold & Hoover, 2000), and victimization rates from 9 percent (Juvonen, Graham, & Schuster, 2003) to 90 percent (Hoover, Oliver, & Thompson, 1993). Thus, for the most part prevalence estimates from small-scale studies differ widely depending on study methods, measurement, and samples, and should be considered (at best) weak approximations of the overall prevalence rate of bullying and victimization for American students. Without consistency in methods and measurement, even studies involving similar populations of students should be interpreted cautiously if changes or variations in prevalence rates are the focus of the investigation.

Solberg and Olweus (2003) discuss the difficulty of establishing reliable prevalence rates for bullying and victimization with respect to the definition of "prevalence" as used in epidemiology. Prevalence, in this sense, denotes the number of individuals with a disease or condition at a specific point in time or within a specific period of time, relative to the total number of individuals in the population who are at risk of or exposed to the condition (Solberg & Olweus, p. 239). Thus, a period prevalence estimate for bullying/victimization "refers to the proportion or percentage of students in a school or other meaningful unit who have been exposed to bullying/victimizing behavior by other students with some defined frequency within a specified time period in the group of interest" (p. 240). Solberg and Olweus contend that the wide range in estimates of bullying prevalence rates is owing to a lack of adherence to this epidemiological definition of prevalence, rather than to meaningful differences in rates of bullying/victimization behaviors. As an addendum to the Solberg/Olweus definition of a bullying prevalence estimate, it would be worthwhile to add further description of the population being studied. Prevalence rates vary considerably for younger and older students, and it would be useful to establish separate rates for elementary, middle, and high school age intervals.
Prevalence in the United States

As discussed in the previous section, the variation among small study prevalence rates is substantial. For example, Schwartz, Dodge, Pettit, and Bates (1997), in an assessment of 198 elementary-school-aged boys, classified 19 percent of the sample as victims and 17 percent as bullies. A much larger victim rate was reported by Hoover et al. (1992), who found that 77 percent of secondary school students had been victims of bullying at school. In addition to the difference in respondent ages as measured in the Schwartz and Hoover studies, the self-report survey item used by Hoover and colleagues was "Have you ever been bullied during your school years?", while Schwartz et al. assessed prevalence rates via peer nomination. Even for studies that use similar methods of measurement and comparable age ranges of students, results are inconsistent across populations. For example, in a study of 154 fifth-grade students in rural Georgia, 23 percent of respondents self-reported being bullied during the past couple of months (Pellegrini, Bartini, & Brooks, 1999). Much higher rates were reported by Stockdale, Hangaduambo, Duys, Larson, and Sarvela (2002), in a rural Illinois sample of 739 children in grades 4 to 6. Stockdale et al. found that 76 percent of students reported having been verbally bullied, while 66 percent reported they had been physically bullied at least once during the previous week.

In April, 2001, Tonja Nansel and colleagues published the results of a nationally representative survey conducted in 1998 of 15,686 public and private school students in grades six to ten. The survey was part of the World Health Organization's Health Behavior in School-Aged Children (HBSC) study, a cross-national investigation of bullying in 30 countries. Nansel et al. found that 30 percent of U.S. students were bullies, victims of bullies, or both during the current school term. More specifically, over 10 percent of the students reported that they sometimes bullied others, and 9 percent acknowledged that they bullied other students at least once a week or more – indicating that nearly 20 percent of the students bullied others "sometimes" or more often. With respect to victimization, more than 8 percent of respondents reported they were sometimes the victims of bullying, and nearly the same percentage said they were bullied at least once a week. Thus, approximately 17 percent of the sample were
victimized "sometimes" or more frequently during the school term. Additionally, 6 percent of the sample reported having bullied others and having been victims of bullying themselves.

In response to Nansel et al.'s study, Greif, Furlong, and Morrison (2003) took issue with the HBSC survey instrument's operational definition of bullying, contending that the specific nature of bullying was not captured by the survey. Greif and colleagues argue that the definition of bullying used does not sufficiently address the component of a recurring pattern of interactions, in other words, does not adequately state that the behavior must be repeated over time. Therefore, Greif et al. contend that estimates in the Nansel study may have been upwardly biased.

In another large-scale study, DeVoe, Kaffenberger, and Chandler (2005) analyzed data from 8,374 public and private school students aged 12 to 18 years who took part in the 2001 School Crime Supplement (SCS) to the National Crime Victimization Survey. The SCS is a nationally representative sample of students aged 12 to 18 and has been administered six times since 1989. Similar to the findings reported by Nansel et al., DeVoe and colleagues found that 14 percent of respondents had been bullied at school in the previous six months. Indirect bullying such as social exclusion was the most common type of victimization (7 percent), while 3 percent of students reported being victimized by direct or physical bullying. Roughly 5 percent of students reported having been bullied both directly and indirectly.

To summarize the discussion of U.S. prevalence rates, (1) there is wide variability in rates across studies; and (2) prevalence rates are affected by multiple factors, including population characteristics, type of measurement, and time frame of measurement. As such, comparisons across studies are of limited value unless equivalent assessment methods are used on similar populations. Given the wide variation in rates of bullying behavior as reported in small-scale studies, the most accurate and most broadly relevant prevalence rates for bullying in the United States are large-scale studies such as those conducted by Nansel et al. (2001) and DeVoe et al. (2005).
Prevalence in Other Countries

Bullying is a pervasive problem worldwide. Not surprisingly, prevalence rates reported in the literature vary between cultures, and between studies within cultures. The validity of cross-cultural comparisons is contingent on the way in which bullying is operationally defined within studies and across cultural settings (Griffin & Gross, 2004; Hay, Payne, & Chadwick, 2004). Differences within U.S. studies underscore that assuming a consistent definition for the term "bullying" is unwarranted even within a single country, which suggests it is even less judicious to assume the term is similarly defined in international research.

Language differences resulting from a non-universal vocabulary is one of the reasons for the difficulty in cross-cultural comparisons (see Arora, 1996; Craig et al., 2000; Kalliotis, 2000). For example, the common translation for bullying in Japanese is "ijime," however some Japanese researchers argue that ijime is not directly comparable to bullying as Western researchers use the term (e.g., Naito & Gielen, 2005; Taki, 2001). Similar translational concerns have been raised for bullying-type behaviors in Korea, known as "wang-ta" (Smith, 2004). Furthermore, languages such as Greek do not have a specific word for bullies, in which the definition "someone who has no discipline and order" has been used for research purposes (Kalliotis, 2000). A multi-national study by Smith and colleagues (2002) in 14 countries addressed the problem of differing cultural definitions of the term bullying. Using cartoons depicting a range of social situations to investigate the meanings given to various terms, the researchers found that while 8-year-olds were only able to discriminate between non-aggressive and aggressive social situations, 14-year-olds were able to discriminate between fighting, physical bullying, verbal bullying and social exclusion. The authors conclude that cultural nuances exist within each language in regard to the terms used for bullying behaviors, and that these should be considered carefully when conducting cross-national comparisons.

Cultural differences in behavior norms are also relevant, particularly for less obvious forms of bullying and for types of bullying that may be interpreted differently across cultures. For example, some researchers claim that laughing at or making fun of others is more common in Italy than in Northern Europe, and that Italian children consider this behavior less
harmful than do children from other European countries (Menesini & Fonzi, 1997, as reported in Gini, 2004). Peer conflicts may also be perceived as more detrimental in some cultures than in others; one study found that conflicts among peers lead to a break in friendship more frequently among Canadian children than among Italian children (Schneider & Fonzi, 1996, as reported in Gini, 2004). Furthermore, the racial/ethnic/religious homogeneity of students in different countries may impact rates and types of bullying. For example, Stein (1999) suggests that the generalizability of Olweus's commonly cited findings may be limited because the populations in Sweden and Norway have little diversity in race, ethnicity, language, or religion. Another cultural difference noted by Stein is that many Western European countries have a standardized, national education curriculum, which allows for easier comparisons and implementation of anti-bullying policies across schools than for countries like the U.S. and Canada which do not have a standardized curriculum.

**Relevance for American Populations**

Definitional issues aside, the relevance of bullying research to the U.S. is not always clear. For example, much of the Japanese anthology of research appears to focus on bullying as an outcome of a competitive academic climate (e.g., Hoover & Hazler, 1991), while findings on bullying in Israel and some Eastern European countries may be influenced by ethnic group conflicts specific to residents of those countries. To the extent that cultural and definitional differences between comparison countries and the United States are minimal (such as Canada, the United Kingdom, Australia, and New Zealand), findings from international research are beneficial for developing a more comprehensive understanding of the bullying phenomenon.

As mentioned previously, research from several other countries has a longer history than research in the U.S., and much of the seminal work on bullying was conducted with non-American populations. In his 20 years of research on school bullying, Dan Olweus has primarily reported on the results of four empirical studies based in Scandinavia. These include:

1. a cross-sectional study conducted in 1983 of 17,000 boys and girls in grades 3 to 9 from 60 schools in three Swedish cities (e.g., see Olweus, 1991);
(2) a cross-sectional study conducted in 1983/1984 of 130,000 boys and girls in grades 2 to 9 from 715 schools in Norway (e.g., see Olweus, 1993b); 

(3) a longitudinal study of the effects of the Olweus Bully Prevention Program, initiated in 1983-1985 and consisting of four cohorts of 2,500 Norwegian boys and girls in grades 4 to 7, along with data from 300 to 400 teachers and principals and 1,000 parents (e.g., see Olweus, 1993b); and 

(4) a longitudinal study that began in 1973 and involves roughly 900 boys originally in grades 6-8 from Stockholm, Sweden (e.g., see Olweus, 1978, 1994).

Olweus (1994) contends that 15 percent of children are involved in bully/victim interactions at some point during their elementary and junior high school years. He estimates that approximately 6-7 percent of children have bullied others, and 2 percent of bullies have harassed others at least once a week. Olweus finds that 9 percent of children are victims of bullying, with 3 percent being victimized at least once per week. Furthermore, about 1.6 percent of students can be categorized as both victims and bullies (see Olweus, 1978, 1979, 1980, 1984, 1991, 1992, 1993c, 1993d). Olweus's findings are widely cited in the bullying literature, but readers should bear in mind that these findings are based on Scandinavian populations, and prevalence estimates are not necessarily generalizable to the United States (or other dissimilar populations).

Research from abroad suggests that many countries have higher rates of bullying and victimization that does the United States. Nansel, Craig, Overpeck, Saluja, and Ruan (2004) compared the results from self-report surveys obtained from nationally representative samples of students in 25 countries, and found substantial variation in the rate of involvement in bullying (which ranged from 9 to 54 percent of youth). For example, Rigby and Barnes (2002), reporting on an Australian survey of over 38,000 primary and secondary school students, found a past year victimization rate of 50 percent, and a past week victimization rate of 17 percent. In a study of 238 Italian middle school students, 30 percent of students reported

---

5 Farrington (1993) notes that the prevalence estimates of victimization from the Norway study are based on 83,330 children in 404 schools.
6 This study is often referred to as the Bergen study.
being victimized "sometimes" or more often in the past three months, while 15 percent reported being bullied at least once a week (Baldry & Farrington, 1999). Even higher rates were found in a British study of 4,700 children aged 11 to 16 years, with 75 percent of students reporting victimization by physical bullying during the past school year (Glover et al., 2000).

Research from Canada is the most clearly generalizable to the United States, owing to the similarity of cultures. In a study of 211 students aged 4 to 14 from 22 classrooms, Charach, Pepler, and Ziegler (1995) found that 20 percent of students reported being bullied more than once or twice during the term, while 15 percent of students reported having frequently engaged in bullying others. Looking at the frequency of involvement rather than overall prevalence rates, Pepler and Craig (2000) estimate that 75 to 80 percent of Canadian children are relatively uninvolved in bullying or victimization (although they may still be negatively affected by its occurrence), 10 to 15 percent of children are occasionally involved, and 5 to 10 percent are involved more than twice a week. Similarly, Craig and Pepler (1997), in a study using naturalistic observations of behavior, reported that bullying occurs once every 7 minutes on school playgrounds. Furthermore, results are available from the nationally representative Canadian HBSC survey, the same survey used in the U.S. by Nansel et al. (2001). Forty-one percent of male and 31 percent of female respondents reported bullying others in the past two months, while 40 percent of male and 43 percent of female respondents indicated they had been a victim of bullying in the past two months (Craig & Harel, 2004). These rates are higher than the findings reported by Nansel and colleagues in the United States.

In summary, prevalence rates of bullying and victimization vary substantially among countries, due in part to study methodology, definitional variation for the term "bullying," and cultural differences in behavior. While interesting that the phenomenon is prevalent worldwide, U.S. researchers should consider the impact of cultural differences when seeking to generalize international study results to American schoolchildren.
MEASUREMENT

As touched on in the previous section, the influence of measurement type on the outcomes obtained from studies on bullying can be considerable. A figure cited in a report such as "25 percent of students are bullied" is vague. Does this rate include all school-aged children from ages 5 to 18? Does it include all forms of bullying – for example, direct physical bullying as well as indirect, anonymous insults or threats posted on a web page? Does it refer to lifetime experiences with bullying or past week experiences? The following section discusses the impact of the time frame of assessment, as well as several different ways in which bullying and victimization are commonly measured.

Time Frame of Assessment

The time frame of the assessment refers to the specific period of time during which bullying is measured, for example, past week, past month, past semester, past year, lifetime, and so on. Prevalence rates of bullying and victimization as well as the frequency of the incidents (e.g., being harassed twice a week versus once a semester) depend on the specific time periods denoted in the definitions used, and on the timing of the assessments themselves. If a questionnaire does not provide respondents with a specific time frame for which to report on their experiences with bullying and victimization, each respondent must develop his or her own interpretation of the questions (Griffin & Gross, 2004). Even when time frames are provided in an instrument, the manner in which questions are phrased and the time frame inquired about may influence the reported rates of bullying and victimization. For example, some researchers claim that when the time frame is short (e.g., past week versus past semester), lower prevalence rates of bullying and victimization are found (e.g., Craig & Pepler, 2003). Others propose that since longer time frames are more dependent on long-term memory, bias is introduced because recall of incidents from five years prior may not be as clear as recall of incidents from five weeks prior (e.g., Hamby & Finkelhor, 2001). Furthermore, Farrington (1993) suggests that even those studies that attempt transparency with respect to time frame may be misleading. For example, some questionnaires refer to bullying "this term,"
yet school terms vary in length and may cover a period from one to four months; results are also dependent upon the point in the school term during which the questionnaire is administered (e.g., beginning, mid, or end of the term). For these reasons, comparing prevalence rates across studies using different time frames is ill-advised.

Types of Measures

There are several types of assessments commonly used to measure bullying, these include: self-report questionnaires, individual interviews, focus groups, teacher reports, peer nominations, and classroom/playground naturalistic observations. Each of these measures is described in more detail in the following section. In addition, the potential for the use of biological measures is discussed.

Although occasionally used as a means to measure bullying, this section does not include a discussion of parent reports. Parent reports are infrequently used, but may be a valuable source of information for very young children because parents often spend many hours of the day with their children and observe them in a range of social situations (e.g., see Ladd & Kochenderfer-Ladd, 2002; Zimmerman, Glew, Christakis, & Katon, 2005). Parent assessments are less useful for school-aged children, however, since parents usually do not observe their children at school. Although parent reports have been used in a few studies of older children (e.g., Weisz, et al., 1993), in general they are not recommended since parents tend to be largely unaware of the amount of bullying and victimization experienced by older children and teens. Furthermore, parent reports are particularly likely to be impacted by social desirability bias, with some parents being reluctant to admit that their children engage in aggressive bullying behavior or are unpopular and picked on by classmates (e.g., LeBlanc, Sautter, & Dore, 2006).

a) Self-Report Questionnaires

By far the most commonly used method for identifying both bullies and victims are self-report questionnaires; these have been used extensively by researchers worldwide (e.g,
Boulton et al., 2002; Olweus, 1994; Rigby & Slee, 1999; Salmivalli, Kaukiainen, Kaistaniemi, & Lagerspetz, 1999). In a self-report, students fill out a questionnaire regarding their experiences with bullying and victimization; questionnaires can be administered to a group of students in a classroom, sent by mail to a student’s home address, completed via e-mail, or completed online. Many researchers believe self-reports are the best way to identify victims, since much bullying, especially relational bullying, is covert and is not witnessed by teachers, parents, or peers (e.g., Dess, 2001). Thus, victims themselves are the best and most valid source of information regarding the various types of bullying or victimization they experience in a variety of settings (Ladd & Kochenderfer-Ladd, 2002). Self-report measures can be implemented quickly, can yield a large amount of data, can be filled out anonymously, require minimal researcher involvement to administer, and are relatively inexpensive. Self-reports can generate information regarding the frequency of bullying and victimization, characteristics of perpetrators and victims, the types of bullying that occur, the circumstances leading up to and following the incidents, and the location in which bullying takes place most frequently. Questions can also assess the emotional, mental, and physical responses to bullying as experienced by victims, perpetrators, and bystanders.

Self-reports are typically written questionnaires, which requires the questions to be presented in a simple and straightforward manner to ensure adequate comprehension by respondents. In some cases teachers or researchers first discuss the definition of bullying along with examples of behaviors to ensure student understanding (e.g., Mynard & Joseph, 2000), or, for younger students, read the questions aloud. Betlem (2001) suggests that providing several examples is important for younger children, who are usually in the Piagetan "concrete operational" stage of cognitive development (Piaget, 1970), and as such are unlikely to consider any behaviors to be bullying beyond the specific examples provided in the questionnaire. Pepler and Craig (1998) question the use of self-report measures with young children, warning that these children often do not differentiate their behaviors from those of their peers, and lack the language capabilities to define more complex relationships and interactions.
Limitations of self-reports

A related concern with self-reports is whether or not a clear definition of bullying and victimization should be provided in the survey instrument. Self-report responses to a general statement such as "I bully others" may have questionable validity if some respondents do not consider their actions to be bullying. Solberg and Olweus (2003) suggest that the provision of a definition reduces subjective interpretation and variability in construct validity across studies, while Espelage and colleagues argue that providing such a definition may "prime" a student and affect the honesty of responses (Espelage, Bosworth, & Simon, 2001; Espelage & Swearer, 2003).

A frequent criticism of self-reports is that they may lead to underreporting of bullying and victimization. Some researchers claim that children have a reluctance to reveal victimization (e.g., Griffin & Gross, 2004; Ladd & Kochenderfer-Ladd, 2002; Neary & Joseph, 1994), while others claim that children tend to underreport their own bullying behavior either because of social desirability effects, or by simply failing to recognize their behavior as bullying (e.g., Griffin & Gross, 2004; Monks, Smith, & Swettenham, 2003; Pellegrini & Bartini, 2000b). While self-report questionnaires tend to be confidential or anonymous, recent research suggests that anonymity may not be necessary for the purposes of gaining valid responses. Chan, Myron, and Crawshaw (2005) tested the issue in a sample of 562 students in grades one through eight at two Toronto schools. The authors found that respondents in the group required to write their names on the questionnaires did not differ from the anonymous group with respect to their reports of bullying and victimization. Although further research is needed before definitive conclusions are drawn, such findings suggest that self-reports of bullying and bully victimization may not require anonymous questionnaires, a fact that would ease the implementation of longitudinal research and would expand the potential for linking student bully/victimization data with other information such as academic outcomes and school disciplinary reports.

Another concern with self-report data is the subjective nature of victimization. Research using multiple informants points to a subset of children who report being victimized
by bullies, but who are not recognized as victims by their peers or teachers (Graham & Juvonen, 1998; Labash, 2003; Perry, Kusel, & Perry, 1988; Salmivalli, 2002). Conversely, other children do not self-report being victimized, yet are listed as victims of bullying in peer nominations (Graham & Juvonen, 1998). This latter finding segues to research that finds some children admit they are victims of bullying but deny being negatively affected by the harassment. For example, Ladd and Ladd (2001) found that over three-quarters of the students in their sample disclosed having experienced victimization, yet only 14 percent reported that this affected them negatively (see also Holt & Espelage, 2003). The issue of subjective resiliency in the face of victimization is an important topic that has received little research focus to date.

b) Interviews and Focus Groups

Other methods for assessing bullying experiences involve qualitative assessments using interviews and focus groups. Valuable insight may be obtained from students willing to speak openly and candidly about their experiences as bullies or victims, including details and follow-up that may not be adequately addressed in written questionnaires. As interviews and focus groups rely on verbal articulation rather than on literacy skills, they may be more suited to students who prefer verbal expression or who have difficulty reading or writing. Conversations also allow the researcher to assess respondents' levels of sincerity and engagement in the research (Rigby, 2002c), and they allow the interviewer to assess and respond to non-verbal cues such as discomfort with particular questions (Hamby & Finkelhor, 2001).

Limitations of interviews and focus groups

Interviews and focus groups have several limitations, such as being poor choices for assessing prevalence rates and for obtaining information that is generalizable beyond the (normally) small samples of children involved in the research. Furthermore, the reliability of findings may be questionable, as different interviewers may elicit different responses from study participants, owing to interviewer personal characteristics, experiences, and biases. A less experienced interviewer could find substantially different outcomes from the same sample of
students than an experienced interviewer who is able to effectively use probing techniques to increase the quality and depth of subject responses.

Getting students to open up about their experiences as victims and bullies may also be a challenge if students are reluctant to portray themselves negatively (i.e., as aggressive or submissive) in the eyes of the interviewer or in the eyes of their peers in the context of a focus group (e.g., Phoenix, Frosh, & Pattman, 2003; Rigby, 2002c). For example, a study by Rigby (1995) found that only a minority of students aged 12 to 16 years were willing to discuss victimization experiences in a group context, despite relatively high rates of reported bullying. Crothers and Levinson (2004) suggest that this can be minimized by having interviews conducted by personnel outside of the school system. However, a study of different data collection types among middle and junior high school students found only a 50 percent correspondence rate for admission of bullying in interviews among those respondents who admitted to bullying on a self-report questionnaire (Ahmad & Smith, 1990, as reported in Ross, 2002). If the issue of lack of disclosure or particularly sensitive questions is a serious concern, computer-assisted self-interviewing (CASI) can be used. This technology allows interviewers to turn a portion of the interview over to respondents themselves, so that respondents can read or listen to questions read aloud via recording and provide their answers in private. Although expensive in terms of equipment and set up, some studies have found that CASI results in an increase in the reporting of sensitive information compared with regular interviews or traditional pencil-and-paper self-report questionnaires (e.g., Turner et al., 1998).

c) Teacher Reports

Teacher report measures of bullying typically involve teachers' observations of student behaviors, or their ratings of individual students' disruptive and aggressive behaviors in the classroom, social skills, and academic competence. In other versions of teacher reports, the teacher compares the behavior of each child with the average behavior of students in the classroom or with a group of similar students. In yet another method, teachers are given a list
of student names, and are asked to identify those who are bullies or victims, or to identify those children who match specific behavioral descriptors (Crothers & Levinson, 2004).

Mostly useful for the assessments of primary-school children (Espelage & Swearer, 2003; Leff, Kupersmidt, Patterson, & Power, 1999; Smith, 2004), teacher questionnaires are valuable in offering perspectives on child behavior that are independent of student friendship dynamics, yet retain a uniquely close and personal vantage point. Some researchers claim that adults are more sophisticated judges of social distinctions among children than are children themselves (LeBlanc et al., 2006). Rigby (2002c) stresses that teacher ratings of secondary school students are unreliable, because students are normally taught by multiple teachers each day and are not under nearly as close a level of supervision as are primary students. One of the major advantages to teacher rating scales is that they are more confidential and less intrusive to students than are self-reports and peer nominations. Teacher report forms are a relatively quick and simple way to gather data, are low in cost, and enable ready comparison of responses among teachers for ratings of the same children (Lagerspetz, Bjorkqvist, Berts, & King, 1982; Rigby, 2002c).

Limitations of teacher reports

Some researchers believe that teachers may vastly underestimate the amount of bullying that takes place at school (e.g., Smith, 2004; Smith & Sharp, 1994; Stockdale et al., 2002). For example, O’Moore and Hillery (1989) found that teachers were able to identify only 24 percent of those students who self-reported themselves as bullies. Reasons for this discrepancy may be that teachers vary in their attentiveness to or awareness of social nuances (e.g., their ability to discriminate between bullying/teasing and play; see Leff et al., 1999). Some research suggests that teachers consider physical bullying to be more serious than verbal or relational bullying (e.g., Hazler et al., 2001), so bullies using non-physical methods of harassment may be more likely to avoid being labeled as a bully by their teacher. Teachers may also be affected by classroom disruptions, and they have immediate classroom instructional responsibilities that prevent them from noticing child behavior (Griffin & Gross, 2004). Much
bullying behavior occurs out of teacher sight, and many of the indirect bullying methods such as rumor-spreading and covert teasing are easy to conceal from teachers.

Furthermore, teacher ratings may reflect teacher biases. Teachers' responses are the result of their own experiences with students in specific settings, such as in a fifth-grade classroom, and bias may occur if teachers are asked to consider a wider range of settings or age levels (Pellegrini & Bartini, 2000b). For example, Leff et al. (1999) examined the ability of 61 teachers to identify peer-reported bullies and victims in a sample of 1,139 students, finding that both African-American and Caucasian teachers were more accurate in their identification of African-American bullies than they were of Caucasian bullies. Teachers were also more accurate in their identification of bullies and victims at the elementary school level than at the middle school level. In addition, teacher reports may be affected by relational biases such as halo effects, since they come to know children fairly well over the course of an academic year, and may interpret student behaviors in light of pre-existing expectations (Ladd & Kochenderfer-Ladd, 2002).

d) Peer Nominations and Peer Relations Surveys

Another way to collect information on bullying is by examining a child's social status among their peers. Peer nomination methods may include, for example, asking students to answer the question "Who in your class is picked on or teased by other kids?", having children match other students' photographs with behavioral descriptors such as "like/dislike" or "bully/non-bully" (e.g., see Bowers, Smith, & Binney, 1994), or requesting students to list three peers who fit various descriptions (such as peers who are disliked, peers who are their friends, peers who are teased, and peers who bully others; Solberg & Olweus, 2003). Peer rating methods are similar; children may be asked to rate every student in the class on the frequency of direct and indirect bullying behaviors, e.g., "Susie Smith threatens other students: never, sometimes, or a lot" (Solberg & Olweus, 2003). In another version of a peer assessment, Dodge and colleagues (1990) asked children to give "liking" and "disliking" nominations for their
peers, and then summed the nominations for each child. In yet another example, Boulton and Smith (1994) gathered peer nominations through interviews with each child in the classroom.

Some research suggests that children are proficient at identifying both bullies and victims in the classroom (e.g., Pakaslahti & Keltikangas-Jarvinen, 2000), while other studies find that peers are good at identifying bullies but are less reliable at identifying victims (Monks et al., 2003). For the most part, peers may be more aware of covert bullying behavior than are teachers (Ladd & Kochenderfer-Ladd, 2002; Ross, 2002). Because measures among students are usually completed by multiple informants in the classroom or even the entire class of students who have been interacting with their peers for quite some time, this type of assessment has been praised for its accuracy, reliability, and validity (e.g., Rigby, 2002c).

**Limitations of peer nominations**

Anonymity is a problem with peer nominations, and social pressures or fear may inhibit students from responding honestly, or may influence the nomination of certain peers over others. Nominations may also be influenced by current or recent social conflicts among students, leading to nominations that are not reliable indicators of student behaviors, for example a student nominating a peer whom he or she dislikes, or nominating a peer as a joke. Furthermore, peers may also be unaware of bullying behaviors they are not involved with personally. These factors lead to the possibility of underreporting (Cole, Cornell, & Sheras, 2006). However, Cole et al. (2006) also contend that peer nominations may identify more bullies than do self-reports, suggesting that peer reports do not suffer as much from the problem of lack of disclosure. Another concern with peer reports is the necessity for informed consent procedures, which may preclude the participation of all children in a classroom or school, leading to missing data that biases results (Griffin & Gross, 2004). In addition, Rigby (2002c) suggests that ethical problems may be more common with peer nominations, since respondents are being asked to judge and report negatively on their fellow students. Some research indicates that peer nominations are not appropriate for very young children; a study by Goodman, Stormshak, and Dishion (2001) found that peer nominations were reliable and valid measures of peer victimization for students in the fifth-grade, as evidenced by test-retest
reliability and convergent validity with teacher reports. The converse was found for peer
nominations of students in the first-grade, with a lack of reliability and validity evidenced by
unstable test-retest measures and no consistent correlations with teacher reports of
victimization.

e) Direct Observations

A less frequently used method of assessing bullying is through direct observation of
children's interactions. This approach presents some methodological challenges but may be a
rich source of information regarding the social context of bullying behavior. Direct
observations, also called "naturalistic observations", can be relatively unstructured.
Observations can also be structured to gather information related to a specific issue in the
bully/victim relationship, such as bullies' use of derogatory language in reference to a victim’s
ethnic background or sexual orientation.

Observational methods employ trained observers who record aggressive, cooperative,
and submissive behaviors for selected children or, less obtrusively, use video cameras or
remote microphones. The general procedure involves researchers selecting a time during
which bullying is to be assessed in a specific location such as the playground, classroom,
cafeteria, or school bus, then making a series of repeat observations which last several minutes
each (e.g., Crick, Ostrov, Appleyard, Jansen, & Casas, 2004). For example, Craig et al. (2000)
used video camera recordings to compare the frequency, duration, and nature of direct and
indirect bullying on playgrounds and in classrooms in a sample of 34 children in grades 1
through 6. Children involved in the study wore small remote microphones and a pocket-sized
transmitter in order to record their conversations with those around them. The advantages of
direct observations are that, if done unobtrusively, such observations may provide unbiased
and objective assessments of students' behavior, they are inexpensive, and they are high in
external validity because they are carried out in naturalistic settings – including the playground
and cafeteria or other settings that tend to be less supervised than the classroom (e.g., Pepler &
Craig, 1995). Observations also help diminish the problem of respondents tending towards
socially desirable responses in interviews and non-anonymous self-report questionnaires (Rigby, 2002c). For example, Pepler and Craig (1995) cite that in their own research, self-report survey results suggest that fewer girls than boys admit to bullying, while naturalistic observations find that girls bully at the same rates as do boys. Finally, video or audio recordings can help illuminate underlying bullying processes, by providing a retrospective opportunity for students to view their own behavior and explain their perspective in a particular bullying incident (Betlem, 2001).

**Limitations of direct observations**

The time constraints and the degree of effort required to conduct naturalistic observations are considerable obstacles to the widespread use of such techniques. In addition, it may be difficult to observe peer aggression unobtrusively, such as when observers are watching from close quarters or when children are required to carry microphones (Rigby, 2002c). For observers, the decision about the optimal distance from which to conduct observations is important, as being too close may cause self-consciousness in children and lead to a restriction or change in their interactions (a Hawthorne effect), while being too far may cause observers to miss softly-spoken verbal threats or other hidden behaviors (Pepler & Craig, 1995). Furthermore, even if the subjects' interactions are properly observed and recorded, the incidence of bullying may be underestimated because much bullying is covert and indirect (Atlas & Pepler, 1998; Colvin, Tobin, Beard, Hagan, & Sprague 1998; Olweus, 1993a). This method is particularly limited for identifying indirect bullying techniques among older students who are more likely to try to avoid detection (Pepler & Craig, 1995). Not surprisingly, research suggests that students prefer to bully others in settings away from adult supervision (such as in restrooms and in locker rooms), rather than in the classroom or in other readily observable settings. Thus to the extent that students are aware of the placement of observers or recording equipment, bullying may be displaced to other locales (Crothers & Levinson, 2004).

In addition, some research suggests that observational measures do not correlate well over time, perhaps because of small samples of observations and transient behavior that is
situationally specific (Pellegrini & Bartini, 2000b). One final disadvantage is that direct observations may be considered obtrusive, and obtaining informed consent for student participation may be more onerous than for other types of bullying measures (Pepler & Craig, 1995).

f) Biological Measures

Although biologic markers have yet to be used in mainstream bullying research, recent studies on hormones present promising ideas for future inclusion of such measures in the bully researcher’s repertoire. Certain hormones present in human saliva have been linked to behaviors such as aggression, depression, and stress (Goodyer, Herbert, & Altham, 1998; Granger et al., 2003). Saliva sample collection is straightforward and can be conducted virtually anywhere; new analysis techniques allow saliva to be tested for levels of testosterone, cortisol, and dehydroepiandrosterone (DHEA; Hazler, Carney, & Granger, 2006). Hazler et al. (2006) suggest that these three hormones may be particularly useful in bullying research as they are markers of emotions and symptoms related to bullying and victimization.

Concordance of Measurement Techniques

There are often wide variations in estimates of bullying as a result of these different methods of reporting. A number of studies have triangulated various measurement techniques and report conflicting findings about the correspondence between measures. The concordance between self-reports and peer reports of bullying has been studied the most frequently, and research usually find that the two measures correspond anywhere from low to moderately well (e.g., Monks et al., 2003; Olweus, 1991, Pakaslahti & Keltikangas-Jarvinen, 2000; Perry et al., 1988). Ladd and Kochenderfer-Ladd's (2002) summary of the literature finds correlations ranging in size from .14 to .42, while a meta-analysis by Card (2003) involving 205 studies found a correlation of .37 between peer reports and self-reports of victimization. According to Graham and Juvonen (1998), self-reports and peer reports assess victimization from different perspectives, and as such are linked to different risks for negative outcomes. Specifically, self-
reports represent subjective experiences of being victimized, and are linked to intrapsychological consequences such as depression, anxiety, and low self-esteem. Peer reports of victimization represent group member's judgments of their classmates' reputations, and predict interpersonal consequences such as being liked, disliked, accepted, or rejected by peers.

The correspondence between peer and teacher reports of bullying ranges from moderate (e.g., Coie & Dodge, 1988; McNeilly-Choque, Hart, Robinson, Nelson, & Olson, 1996; Perry et al., 1988) to high (e.g., Ladd & Kochenderfer-Ladd, 2002; Monks et al., 2003; Salmivalli & Nieminen, 2002; Schuster, 1999). For example, a study by Pakaslahti and Keltikangas-Jarvinen (2000) assessed both direct and indirect aggression in a sample of more than 2,000 students and found a higher correlation between peer and teacher reports than between self- and teacher reports and between peer and self-reports. Similarly, Xiao and Matsuda (1998) found that reports of bullying from peers and teachers were significantly more likely to correspond with each other than were reports from peers and self, or reports from teachers and self.

Rates of correspondence for self-reports and teacher reports vary substantially among studies, however research tends to find that these measures do not correlate well (e.g., Monks et al., 2003; Xiao & Matsuda, 1998). For example, Schuster (1999), in a German study of 442 students in grades 5, 7, and 11, found a correlation of .23 for self-reports and teacher reports. In another study, Barone (1997) reported on a survey administered to 847 eighth-graders and 110 counselors, teachers, and administrators in upstate New York. On average, school personnel estimated that 16 percent of the students had been victims of bullies, compared with a rate of 60 percent obtained from the student reports.

Factors Affecting Measurement Correspondence

The correlation among measures may be affected by characteristics such as the age of students under study. For example, Pakaslahti and Keltikangas-Jarvinen (2000) found that the correspondence between teacher and peer reports was higher for younger students than for older students, while the correspondence between self-reports and teacher reports and
between self-reports and peer reports had curvilinear relationships, in other words, lower rates were found for the youngest and the oldest students compared with the mid-aged students. The meta-analysis by Card (2003) also reported that the concordance rates among methods of assessing victimization were smallest for studies of young children (see also Ladd & Kochenderfer-Ladd, 2002).

Another characteristic that appears to affect the correlation between measures is gender. For example, a study by Crick et al. (1997) of 65 preschoolers found a correlation of .32 for teacher- and peer reports of overt aggression in boys, and a correlation of .11 for teacher and peer reports of relational aggression. For girls, the correlation between teacher and peer reports of overt aggression was .31, and the correlation between teacher and peer reports of relational aggression was .42. In other words, the correspondence between teacher and peer report measures of relational aggression was substantially higher for girls than for boys. In another study, Pakaslahti and Keltikangas-Jarvinen (2000) found a higher correlation between peer and teacher reports of aggression for boys than for girls.

The type of bullying being assessed may also affect the correspondence between informants, as may the identification of bullies versus victims. For example, Monks et al. (2003) found that agreement among informants differed for bullies and victims, with concordance highest between peer and self-nominations for victims, and between teacher and peer nominations for bullies. In another study, Cullerton-Sen and Crick (2005) used a sample of fourth-graders and found a moderate correlation between teacher and peer reports of relational victimization ($r = .34$), and a somewhat smaller correlation between teacher and peer reports of physical victimization ($r = .21$).

**Summary and Conclusion on Measurement**

A study’s selection of instrumentation and methodology to assess bullying behavior ultimately influences the type and quality of the data collected, which in turn influences the results and conclusions that are made. Because of the potential unreliability of single informants (e.g., self-report, peer report, or teacher report), and to increase the validity of the
data obtained, whenever possible study designs should incorporate multiple informants, as well as corroborating behavioral information obtained through direct observation or other evidence of behavioral and emotional difficulties (Griffin & Gross, 2004; Ladd & Kochenderfer-Ladd, 2002; LeBlanc et al., 2006; Monks et al., 2003). Depending on the age of the students in the sample and on the type of bullying being assessed, a different combination of measures might be most useful. For younger students, combining teacher and peer ratings or direct observations may be the most effective for identifying bullying, while a combination of self- and peer reports may be best for identifying victims. For older students, teacher reports are less useful, and self-reports are likely the best method for gaining reliable information on both bullying and victimization. In general, the use of multiple informants combined with the use of multiple assessment methods reduces the relative influence of either the respondents or the method of assessment on the validity of the data (Crothers & Levinson, 2004).

The combination of informant reports may have implications for identifying students at particularly high risk for adverse outcomes of bullying and victimization. For example, a study by Graham, Bellmore, and Juvonen (2003) of 350 African-American and 435 Latino sixth-graders found that differences in adjustment difficulties were related to peer versus self-reports of victimization. Subjects who both self-identified as victims and who were nominated by peers as victims experienced the worst outcomes overall. Those students who self-identified as victims but who were not nominated by peers had significantly more difficulties with psychological maladjustment than did those students who were nominated by other students but who did not consider themselves victims.

Similar to the reasons for advocating a concise, universally-used definition of bullying, bullying research would benefit by focusing on a limited selection of survey instruments (Griffin & Gross, 2004). Many studies develop original instruments to assess bullying, which are created for a specific study population, and differences among such measures lead to difficulties in comparing results across studies. Furthermore, many of the non-standardized questionnaires may not provide valid or comprehensive assessments of the bullying phenomenon, thus limiting the opportunity to collect meaningful data. For example, Colvin et
al. (1998) compared several bullying questionnaires and found that relatively few addressed the range of problems thought to be related to bullying.

Future research on instrumentation should give more consideration to child developmental levels. According to Griffin and Gross (2004), the majority of the currently available instruments with adequate bullying content are limited because they are not validated and normed across developmental levels and age groups. For example, instruments found to be reliable with primary school students are not necessarily effective with middle schoolers. Instruments that are valid for boys are not necessarily adequate for girls, and it may be that instruments used across different cultural groups are not consistently well-understood.

Another challenge regarding measurement includes informed consent procedures. While schools and teachers may offer strong support for bullying research, human subjects committees often require active consent on the part of all students and/or parents. Parents who refuse to consent – or, more commonly, fail to consent – restrict the sample accordingly, resulting in the loss of valuable information about bullying. In particular, informed consent is problematic for peer nomination procedures. In such cases, the list of names that children are given from which to make their peer nominations is shortened (because those children without consent are removed from the list), possibly eliminating the most aggressive or victimized students in the classroom and leading to spurious results (Griffin & Gross, 2004).

Finally, the specific methodology and instrumentation employed in a study should be guided by an ex ante decision about how the data will subsequently be used (Crothers & Levinson, 2004). The appropriateness of techniques and instruments for data collection depend on whether the intent of the study is to assess the overall prevalence of bullying in a school, determine rates of cyberbullying, provide a more in-depth assessment of victim outcomes, or examine the potential for effective intervention. Methods and instrumentation should be selected accordingly.
FOUR PARTICIPANT ROLES: BULLIES, VICTIMS, BULLY-VICTIMS, AND BYSTANDERS

The bullying process involves four key roles, and virtually every student has assumed one or more of these roles at some point during their attendance at primary or secondary school. Bullies and victims are the obvious and most widely-researched roles, while bully-victims, also known as "provocative victims," are a category of children and youths who engage in both types of behaviors. Bystanders are those students who observe bullying incidents but who are not directly involved as a bully or victim. In the following sections each of these four roles is described in detail.

I. Victims (Submissive)

Characteristics of Submissive Victims

Most bullying research suggests there are two main types of victims: passive/submissive (the stereotypical image of a target), and provocative (also known as bully-victims; Olweus, 1993c). According to Olweus, submissive victims are more common, and are typically seen as anxious, cautious, quiet, sensitive and insecure. When physically or psychologically attacked by others these children usually respond by withdrawing, crying, displaying other signs of pain and suffering, or capitulating and giving the bully the object or response that is desired. Passive victims have difficulty asserting themselves among other children, seldom discuss their victimization with others, and rarely defend themselves effectively or retaliate against bullies (Griffin & Gross, 2004). Submissive victims have low self-esteem and often blame themselves for the attacks because they are deserving, worthless, or failures (Salmivalli, 1998). They usually lack social skills and have difficulty establishing friendships, leading to isolation and loneliness at school (Olweus, 1993c). In a study by Fox and Boulton (2005) that used self-, peer, and teacher ratings to examine the specific social skills problems that lead to victimization, the authors found that victims of school bullying were likely to:

(1) display a behavioral vulnerability (e.g., looking scared), (2) be non-assertive (e.g., "give in to the bully too easily…."), (3) reward and thus
reinforce the bully’s behavior (e.g., “cries when picked on”), (4) be withdrawn and solitary in their behavior (e.g., “talks very quietly”), and (5) be quite “provocative” (e.g., “annoys other kids…”). (p. 324)

While it is true that children are often picked on for being "different" in any way, it is a myth that physical differences are the main criteria by which bullies identify their victims (Olweus, 1994). Physical characteristics alone do not predict which students will become victims, and traits such as attire, hygiene, freckles, wearing glasses, and foreign or regional speaking accents do not appear to be significant factors for victimization (Olweus, 1993c). In his early research comparing 201 sixth-grade male bullies and victims, Olweus (1978) found that the only physical characteristic significantly associated with victimization was physical weakness, with smaller and weaker students reporting a greater likelihood of victimization. As stated by Olweus (1993c), "By and large, the victims were found to be no more externally deviant (with regard to 14 external characteristics assessed by means of teacher ratings) than a control group of boys not bullied" (p. 31). He concluded that with the exception of weakness, physical characteristics typically are not as significant as personality traits such as insecurity and social immaturity in eliciting bullying.

Despite these findings, Olweus contends that while an external deviation is not likely to be the reason for victim selection; a bully might pick on and make use of external deviations in a bully/victim situation (1993c). Emerging research on children with disabilities does suggest that children who have particular physical disabilities or medical conditions may be more frequent targets of bullying. Such conditions include: stuttering (Davis, Howell & Cooke, 2002; Hugh-Jones & Smith, 1999); visual impairment (Royal National Institute for the Blind, 2000); hearing impairment (Kent, 2003); craniofacial anomalies (Broder, Smith, & Strauss, 2001); bladder exstrophy (Wilson, Christie, & Woodhouse, 2004); hemophilia (Kvist, Kvist, & Rajantie, 1990); epilepsy (Wilde & Haslan, 1996); cerebral palsy (Tentoglou, 2001); diabetes (Storch et al., 2004, 2006a), cancer (Lahteenmaki, Huostila, Hinkka, & Salmi, 1990); hemiplegia (Yude, Goodman, & McConachie, 1998); and eczema (Lewis-Jones, 2006).
Similarly, children with non-visible disabilities or conditions are also bullied at a higher rate than are non-disabled children. For example, Thompson, Whitney and Smith (1994) studied bullying among special needs children enrolled in mainstream classes in three primary and five secondary schools in the UK. They found that 66 percent of the special needs children reported being bullied by their mainstream classmates, compared with a victimization rate of 25 percent for the mainstream students. Furthermore, more students with moderate learning disabilities were bullied than were students with mild learning disabilities. Other conditions that have been linked with bully victimization include Asperger's syndrome (Little, 2002); specific language impairment (Knox & Conti-Ramsden, 2003); dyslexia (Humphrey & Mullins, 2002); attention-deficit hyperactivity disorder (Unnever & Cornell, 2003); and obsessive-compulsive disorder (Storch et al., 2006b).

Research suggests it is not the differences or disabilities themselves that attract bullying, but the associated behavior displayed by the victimized child. For example, in the case of attention-deficit hyperactivity disorder (ADHD), children with ADHD may exhibit poor social skills and frequent inappropriate behavior, which peers find obtrusive (Unnever & Cornell, 2003). With respect to physical disabilities, children may be self-conscious and display characteristics of insecurity or emotional sensitivity, making them appealing targets for bullies. For example, Dawkins (1996) compared bullying experiences in a group of pediatric outpatients with conditions that affected their appearance or gait with a group of pediatric outpatients with conditions that did not present a visible abnormality. Children who experienced the most bullying were male, had fewer friends, were alone during playtime, and required extra help in school. After controlling for those four factors, results did not find that the children with visible abnormalities were more likely to be bullied. (If course, if visible abnormalities are correlated with these predictive factors, they may be indirectly related to bully victimization).

Other research has uncovered a plethora of additional factors linked with childhood bullying, including: a mentally disabled sibling (Gath, 1992); lesbian or gay parents (Sears, 1994); short stature (Sandberg & Michael, 1998; Voss & Mulligan, 2000), overweight (e.g.,
Elgar, Roberts, Moore, & Tudor-Smith, 2005; Griffiths et al., 2006); a primary language other than English at home (Yu, Huang, Schwalberg, Overpeck, & Kogan, 2003); an alternative sexual orientation (Pilkington & D’Augelli, 1995); atypical gender identity (Wilson, Griffin, & Wren, 2005); unconventional dental appearance (DiBiase & Sandler, 2001); glasses or an eye patch (Horwood, Waylen, Herrick, Williams, & Wolke, 2005); and being viewed as physically unattractive, physically well-developed (girls), or unstylish (Shakeshaft et al., 1995; Sweeting & West, 2001).

According to Pilkington and D’Augelli (1995), victimization based on known or assumed sexual orientation is the most frequently occurring form of bias-related violence in the United States. Using a sample of 2,064 students in grades 8-11, the American Association of University Women (2001) found that being called "gay" or "lesbian" would be considered "very upsetting" for 74 percent of boys and 73 percent of girls, and being called gay would be more upsetting to boys than any other form of sexual harassment in schools. In 2005, the Gay, Lesbian and Straight Education Network surveyed 1,732 lesbian, gay, bisexual, and transgender youth in grades K-12 from all 50 states and the District of Columbia, reporting that over 75 percent of respondents frequently or often heard homophobic remarks such as "faggot" or "dyke" at school.

In sum, the conclusions from studies on characteristics associated with victimization are that (a) victims are usually selected for personality traits such as weakness, lack of social skills, over-sensitivity, or low self-esteem, and (b) once a suitable victim is identified, bullies will make use of almost any defining characteristic, be it physical, cognitive, or personal, in their harassment of the victim.

**Expert agreement on victim characteristics**

Expert agreement on the defining characteristics of victims was assessed by Hazler et al. (1997). In this study, a group of 14 bullying experts from around the world was asked to rate the importance of each of 70 characteristics in identifying victims of bullying. The experts were in strong agreement with respect to 21 characteristics of victims, more specifically, at least
70 percent of the 14 respondents agreed that these 21 characteristics are important factors for victimization. See Table 2.

### Table 2 Characteristics of Victims as Rated by Experts

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Believe that they cannot control their environment</td>
</tr>
<tr>
<td>2</td>
<td>Ineffective social skills</td>
</tr>
<tr>
<td>3</td>
<td>Poor interpersonal skills</td>
</tr>
<tr>
<td>4</td>
<td>Less popular than others</td>
</tr>
<tr>
<td>5</td>
<td>Underlying fears of personal inadequacy</td>
</tr>
<tr>
<td>6</td>
<td>Blame themselves for their problems</td>
</tr>
<tr>
<td>7</td>
<td>Are given labels suggesting inadequacy</td>
</tr>
<tr>
<td>8</td>
<td>Isolated socially</td>
</tr>
<tr>
<td>9</td>
<td>Afraid of going to school</td>
</tr>
<tr>
<td>10</td>
<td>Physically younger, smaller and weaker than peers</td>
</tr>
<tr>
<td>11</td>
<td>Limited skills for gaining success and acceptance</td>
</tr>
<tr>
<td>12</td>
<td>Run out of communication capabilities during high stress incidents</td>
</tr>
<tr>
<td>13</td>
<td>Poor self-concept</td>
</tr>
<tr>
<td>14</td>
<td>Physical mannerisms associated with depression</td>
</tr>
<tr>
<td>15</td>
<td>Frequent feelings of personal inadequacy</td>
</tr>
<tr>
<td>16</td>
<td>Perform self-destructive actions</td>
</tr>
<tr>
<td>17</td>
<td>Believe others are more capable of handling various situations</td>
</tr>
<tr>
<td>18</td>
<td>Difficulty relating to peers</td>
</tr>
<tr>
<td>19</td>
<td>Family members are overly-involved in their decisions and activities</td>
</tr>
<tr>
<td>20</td>
<td>Perceived progressive failures cause them to put forth less effort with each presenting opportunity</td>
</tr>
<tr>
<td>21</td>
<td>Feel external factors have more of an impact on them than internal control</td>
</tr>
</tbody>
</table>

*Adapted from Hazler et al. (1997), p. 9*

### Stability of Victim Status

Many studies find that rates of victimization decrease throughout the elementary school, middle school, and high school years (e.g., Baldry & Farrington, 2005; Boulton & Underwood, 1992; DeVoe, Kaffenberger & Chandler, 2005; Melton et al., 1998; Smith, Madsen, & Moody, 1999; Whitney & Smith, 1993). Nansel and colleagues (2001) found that one-quarter of sixth-graders reported regular to frequent victimization during the school term, while less than one-tenth of tenth-graders reported similar experiences. An age decline was also found in a four-year longitudinal study of nearly 3,500 Australian students (reported by Rigby, 2002c). For both boys and girls aged 8 to 11 years, reports of weekly bullying during the
current school year decreased annually (with the exception of a constant victimization rate of 33 percent for girls aged 8 and 9). Some research proposes a curvilinear trend for victimization, with an increase through the elementary school years, a peak in middle school/junior high (ages 9 to 15), and a decline throughout high school (e.g., Olweus, 1993c; Pellegrini & Long, 2002). For example, Eslea and Rees (2001) conducted a retrospective study of over 400 adults aged 18 to 55 years to assess the ages at which they had experienced bullying at school. The authors found evidence of a normal distribution in reports of bullying over the childhood years, with a peak in rates between the ages of 11 and 13 years. No significant differences were found in recollections of bullying for males versus females, or for younger adults (age 18-21 years) versus mature adults (age 22-55 years).

Several explanations have been proposed for the age decline in victimization. First is the logical argument that since children are not likely to be bullied by peers much younger than themselves (e.g., a seventh-grade student is unlikely to be bullied by a third-grade student), younger children simply have a larger pool of potential bullies at school (e.g., Hay, Payne, & Chadwick, 2004; Kumpulainen, Rasanen, & Henttonen, 1999; Smith, Madsen, & Moody, 1999). Younger children are also less socially competent than are older children, and may not yet have acquired the social and assertiveness skills to effectively combat bullying incidents (e.g., Hay, Payne, & Chadwick, 2004; see also Olweus, 1994). Another proposed explanation for the higher rates of reported bullying among younger children may be differences in comprehension of the term "bullying," with young children believing any objectionable interpersonal interaction to be bullying (Salmivalli, 2002; Smith et al., 2002; Smith, Madsen, & Moody, 1999). Furthermore, Camodeca, Goossens, Terwogt, and Schuengel (2002) suggest that since the experience of victimization evokes strong, unpleasant feelings, many young victims may be motivated to change their behavior in order to avoid future bullying, resulting in decreased rates of future victimization.

Another explanation for the surge in bullying during middle school is that these years are also beset with puberty. This is a particularly challenging time of human development, burdened with rapid physical and psychological changes (Hodges & Perry, 1996; Stillion,
1994), including moodiness induced by hormonal fluctuations (Stillion, 1994). Some researchers contend that the way in which American schools are organized contributes to the incidence of bullying, due to an interaction with child psychosocial and biological developmental levels (Carney & Merrell, 2001). Differences among students in terms of biological maturity are clearly noticeable during this time period, including surges in height and weight, the development of secondary sexual characteristics, and male voice deepening. Weaker boys tend to be picked on by stronger boys, and the combination of adolescents with wide-ranging maturity levels in middle schools may precipitate bullying situations.

Some studies refute the age decline in victimization, including Craig and Pepler (1997); Hoover, Oliver, and Thompson (1993); Johnson et al. (2002); and Slee (1993). For example, a cross-sectional study of 165 American students found no difference in peer-reported victimization rates for children between the ages of 8 and 12 years (Perry, Kusel, & Perry, 1988). Similarly, Johnson et al. (2002), in a study of 523 British children aged 7-11 years, found no significant effects for age. A longitudinal study by Olweus (1977) assessed 201 sixth-grade boys who were rated by peers for bullying and victimization behaviors. Olweus found that a significant number of victims remained victimized at three-year follow-up, and that 65 percent of ninth-grade victimization was predicted by the level of victimization in sixth-grade. In other research, an eight-year longitudinal study of nearly 600 Finnish schoolchildren found that those children characterized as victims at age 8 retained their status at age 16 follow-up (Sourander et al., 2000).

Salmivalli (2002) argues that if there is a true age decline in victimization, it should be confirmed in reports by peers and teachers. In her study of 1,220 children aged 9 to 12 years from 16 schools in Finland, a decrease in self-reported victimization was found, but this decrease was not corroborated by peers or by teachers (whose reports remained steady for children of all ages). Salmivalli suggests that what actually decreased as a function of age was the number of children who self-report being bullied but who do not have the reputation of victim with their peers. This was especially true for children aged 9-10 years, who commonly reported experiencing victimization but for whom peer concurrence was absent. While it may
be that these students truly are bullied and that other students are simply unaware, it is also possible that some students perceive victimization where it does not actually exist.

To summarize, while most studies do find that victimization rates are correlated with age, findings are not consistent across the literature. Victimization does, however, appear to be more common for elementary and middle school students than for high school students.

**Victim Risk and Protective Factors**

No single overarching risk factor for victimization exists, rather, an individual’s personal, familial, and interpersonal characteristics combine to form his or her overall risk level for becoming the future target of a school bully (Smith, 2004).

**Individual factors**

Personal characteristics indicative of internalizing symptoms are frequently linked with bully victimization, including characteristics such as being submissive, cautious, withdrawn, anxious, depressive, and overly sensitive (e.g., Card, 2003; Fekkes, Pijpers, Frederiks, Vogels, & Verloove-Vanhorick 2006a; Goldbaum, Craig, Pepler, & Connolly, 2003; Hodges & Perry, 1999; Olweus, 1993a). For example, Schwartz, Dodge, and Coie (1993) employed observational methods to study play groups of unacquainted 6- and 8-year-old boys. Schwartz and colleagues found that children categorized as victims at the conclusion of their study displayed a different pattern of behavior over the course of the five-day assessment. These children were more submissive than other children, were passive and inflexible in their play activities, and rarely initiated prosocial interactions with their peers. Similarly, Perren and Alsaker (2006) studied 344 kindergarten students and found that victims tended to be submissive, withdrawn, and isolated, and were less cooperative and social compared with children not involved in bullying.

Low self-concept is another major risk factor for bully victimization (e.g., Card, 2003; Salmivalli et al., 1999). For example, Egan and Perry (1998) studied 189 students in grades 3 through 7, collecting data in the fall and the spring of the school year. The authors found that students with low self-regard were more likely to be victimized by their peers, especially in
Egan and Perry suggest that low self-regard may act as a moderator variable, as evidenced by findings that risk factors such as anxiety, lack of social skills, or physical weakness in the Fall term were more likely to lead to Spring term victimization for those children who also had low self-regard.

A lack of social skills, social skills problems, and difficulties engaging in appropriate expression of emotion are additional risk factors for victimization by peers (e.g., Fox & Boulton, 2006; Johnson et al., 2002; Perren, von Wyl, Stadelmann, Burgin, & von Klitzing, 2006). A study by Champion, Vernberg, and Shipman (2003) of 54 early adolescents found that victims of bullying exhibited difficulties in managing confrontation adaptively during interactions with peers. Troy and Stroufe (1987) observed pairs of children at play and reported that victims tend to exacerbate their poor social skills by behaving in an overly needy manner. The victims appeared to long for approval from their peers and continued to (unsuccessfully) attempt interaction even after experiencing rejection or aggression. Socially isolated children are at particular risk for victimization because the presence of friends serves as a buffer from bullies (Limber, 2002), and students rated by peers as popular are less likely to be the victims of bullying (Mynard & Joseph, 1997).

Externalizing behaviors such as aggression, hyperactivity, delinquency, and defiance of teachers have also been linked with bully victimization (e.g., Arseneault et al., 2006; Card, 2003). For example, Perren et al. (2006) studied 153 5-year-old children and concluded that a child's display of behavioral and emotional difficulties coupled with a lack of self-awareness regarding hyperactive/impulsive behaviors elevates their risk for victimization and rejection by peers. In a similar study, Hanish et al. (2004) examined 126 preschool and kindergarten children over the course of an academic year and found that levels of anger in the fall assessment were positively related to spring victimization.

**Familial factors**

A range of family factors have been identified as important predictors of bullying by peers. Harsh, punitive, and inconsistent methods of discipline at home are linked with peer victimization at school (Baldry & Farrington, 2005; Craig et al., 1998; Myron-Wilson, 1999;
Maternal hostility and rejection is a significant predictor of peer victimization, especially for girls (Craig et al., 1998; Finnegan, Hodges, & Perry, 1998; Schwartz et al., 2000), as is negative father-child communication and lack of father involvement in their child's life (Estevez, Musitu, & Herrero, 2005; Flouri & Buchanan, 2002). Children who have been victims of maltreatment such as neglect, physical abuse, or sexual abuse are more likely to be bullied at school (e.g., Duncan, 1999a; Shields & Cicchetti, 2001), and bearing witness to interparental physical violence is another risk factor for victimization. Baldry (2003) studied a sample of 1,059 Italian elementary and middle school students and found that students exposed to interparental physical violence were more likely to be victims of bullying at school, even after controlling for direct child abuse by the father (see also Mohr, 2006). Not all research supports this finding, however, with a study of 112 children aged 6 to 13 years reporting no association between peer victimization and exposure to intimate partner violence (Bauer et al., 2006). Parental marital status may also be a risk factor; Bond, Carlin, Thomas, Rubin, and Patton (2001) found that victims of peer harassment were 1.5 times more likely to have parents who were separated or divorced than to come from intact families.

Research suggests that victims' families often have poor problem-solving abilities, along with interpersonal interaction styles that hamper independence and assertiveness, such as parents being highly demanding and overly-involved in their child's decision making (Hazler, Carney, Green, Powell, & Jolly, 1997; Ladd & Ladd, 1998). Victims tend to be close to their parents, and often have parents characterized by overprotective behavior (e.g., Card, 2003; Olweus, 1993c). For example, Ladd and Ladd (1998) videotaped 197 kindergarten children and their primary caregivers during interactions at home, and then collected peer victimization data on the children as they began kindergarten. Ladd and Ladd found that peer victimization for both boys and girls was associated with parents who displayed high levels of intrusive demandingness and low levels of responsiveness, and that peer victimization in boys was also related to parent-child relationships classified as being intensely close. Olweus (1993d) reported that mothers of chronic victims tended to be controlling and restrictive, and treated their sons as younger than their actual age (see also Beran & Violato, 2004; Myron-Wilson,
The link between victimization and maternal overprotectiveness may be particularly salient for boys; Finnegan et al. (1998), in a study of 184 students with a mean age of 11.7 years, found the link particularly strong for boys who reacted fearfully during conflict with their mothers. Smith and Myron-Wilson (1998) explain this finding by suggesting that overprotective families and overbearing mothers may fail to help children develop assertive social skills and coping strategies, rendering them unable to cope with minor forms of teasing and irritation when interacting with peers. A high level of parental involvement in school activities has also been associated with victimization, with victims being 1.55 times more likely than non-victims to have parents who were heavily involved with their school (Nansel et al., 2001).

**Interpersonal factors**

The most commonly identified risk factor for victimization is a lack of friendships – a risk factor because friends provide important support and protection against bullies (e.g., Hanish, Ryan, Martin, & Fabes, 2005; Hazler et al., 1997; Schwartz, McFayden-Ketchum, & Dodge, 1999). As noted by Smith (2004), "Some children are more at risk (through familial overprotection, shyness, or physical weakness), and this risk of victimization is moderated by both the number and quality of friends they have, and their general standing in the peer group" (p. 100). In a meta-analysis of 205 studies on victim correlates, Card (2003) found the largest effect sizes were for peer group variables, including peer rejection (.38), poor friendship quality (.23), low number of friends (.20) and low peer acceptance (.19).

Studies suggest that children with a higher number of friends are less likely to be victimized by peers (e.g., Fox & Boulton, 2006; Hodges, Malone & Perry, 1997; Mouttapa, Valente, Gallaher, Rohrbach, & Unger, 2004; Schwartz et al., 2000). The quality of friendships is also critical, with low quality friendships serving to increase a child's risk for bullying (e.g., Bollmer, Milich, Harris, & Maras, 2005; Goldbaum et al., 2003). In particular, having a designated "best friend" is a protective factor against peer victimization. For example, Boulton, Trueman, Chau, Whitehand, and Amatya (1999) conducted a short-term longitudinal study with a sample of 158 early adolescents. Peer nominations at Time 1 indicated that students who had a reciprocated best friend in their classroom were less likely to be nominated as
victims than were students without a reciprocated best friend. Students without a best friend at either Time 1 or Time 2 evidenced the largest increase in victimization between Times 1 and 2, while students who had a reciprocated best friend at both Time 1 and Time 2 showed the largest decreases in nominations as victims. Similarly, a longitudinal study of 393 fourth- and fifth-grade children found that having a mutual best friend was a protective factor in victimization, buffering the negative effect of internalizing behaviors such as shyness and anxiety (Hodges, Boivin, Vitaro, & Bukowski, 1999; see also Fox & Boulton, 2006). Pellegrini and Bartini (2000a), in a study using peer nominations, found that both friendships and the size of friend groups are important protective factors against victimization. The authors found that after controlling for baseline rates of victimization, the number of "like most" and reciprocated friendship nominations that children received in the fifth-grade negatively predicted victimization one year later. Pellegrini et al. (1999) suggest that the type of friends a child has affects victimization – having friends who are considered "tough" or a bully themselves decreases the likelihood that a child will become a target for bullying (see also Hodges et al., 1997).

**Reporting and Reactions of Others**

Despite the high prevalence of bullying, substantial numbers of children do not report their victimization to school personnel or to their parents (Bulach, Fulbright, & Williams, 2003; Limber, 2002). For example, a British study found that under a third of regularly bullied secondary students had informed teachers or other school staff about the incidents (Whitney & Smith, 1993), while Smith and Shu (2000) report that fewer than a quarter of bullied students ask an adult for help. Reporting was slightly higher in an American study by Melton and colleagues (1998), who found that 50 percent of students in grades four to six reported their victimization to others. The reluctance to report bullying may reflect a lack of confidence in school personnel's handling of the allegations. For example, Hoover et al. (1992) found that two-thirds of the bullied high school students in their study believed that school personnel responded poorly to bullying incidents at school. Similarly, Harris et al. (2002) reported that
only one-quarter of bullied ninth-grade students felt that things improved after telling someone about the bullying, while 16 percent indicated that things got worse. Newman, Murray, and Lussier (2001) concluded that the more students are motivated by a goal of "getting things back to normal," the more likely they are to approach their teacher for help. Additionally, the less interested victims are in maintaining a friendship with the bully, the more likely they are to seek outside assistance.

Rates of reporting vary by age and gender, with older children and boys being less likely to report victimization (Hunter, Boyle, & Warden, 2004; Melton et al., 1998; Rivers & Smith, 1994; Unnever & Cornell, 2004; Whitney & Smith, 1993). For example, Olweus (1993c) found that 55 percent of primary school victims said they talked with "somebody at home" about their bullying experiences, while only 35 percent of students in secondary and junior high grades reported doing the same. Other factors related to reporting include the chronicity of the victimization, the type of bullying experienced, and the school's perceived level of tolerance for bullying. Unnever and Cornell (2004), in a sample of 2,437 students in six middle schools, found that reporting was more common for students who were subjected to repeat victimization, as well as for students who perceived the school climate to be less tolerant of bullying. With respect to bullying types, a study of 7,000 primary and secondary school students in the U.K. found that boys were more likely than girls to report physical bullying (Rivers & Smith, 1994).

Newman et al. (2001) contend that help-seeking behavior is moderated by gender and popularity, as indicated by a study of 128 third- and fourth-graders from Southern California. The study employed vignettes depicting incidents of peer harassment, and students were asked what they would do if faced with such situations. Fourth-grade girls reported more help seeking intentions than third-grade girls, while the opposite was true for boys. The study also found that self-perception of popularity among peers affected help-seeking, i.e., boys who felt popular said they would seek help relatively frequently, while the opposite was true for girls (those who felt unpopular were more likely to seek help). Newman and Murray (2005), in a similar study, concluded that bullied children may be reluctant to seek help because of
concerns of exacerbating the situation, and that students who are unpopular with their peers are more fearful of retribution from the bully.

While it is often assumed that victims of bullying receive sympathetic reactions from their non-bullying peers, research suggests this is not always the case. For example, Oliver, Hoover, and Hazler (1994) studied 207 Midwestern students and found that a clear majority of students felt that victims were responsible for bringing on the bullying, and many felt that bullying helped the victim by making them tougher. A study by Charach et al. (1995) found that a number of students in grades 5 and 6 considered victims to be "weak," "nerds," and "wimps." When witnessing a bullying situation over 40 percent of the students claimed that they try to help, one-third admitted that they should help but do not, while one-quarter said that bullying was none of their business.

Victim Reactions to Bullying

**Types of reactions**

When children are bullied by their peers they often exhibit a submissive/passive reaction (Olweus, 1978), or react in ways that exacerbate the aggression (e.g., Perry et al., 1988; Schwartz et al., 1993). In an examination of 22 children's books for solutions to bullying, Oliver, Young, and LaSalle (1994) found that 32 percent of the books depicted the main characters engaging in revenge by fighting, while 14 percent depicted coping skills or other resolution strategies for dealing with the bullying behavior.

Submissive or avoidance responses are common; for example, Coie, Dodge, Terry and Wright (1991) found that youths chose submissive responses to bullying in at least half of the situations they encountered. Another study found that 41 percent of middle and junior high school students tried to stay in a group of peers to avoid bullying, while 34 percent of respondents reported staying away from certain locations at school or on school grounds (Chandler, Nolin, & Davies, 1995). Similarly, Smith and Sharp (1994) reported that 67 percent of victims of bullying attempted to avoid the perpetrator, 57 percent tried to stay close to other students, and 20 percent resorted or staying home from school.
More details about the different strategies that children use to cope with bullying have been addressed by several researchers, including Naylor, Cowie and del Rey (2001) in a sample of 324 British secondary school students. The authors found that 86 percent of students told someone such as an adult or a friend about the bullying, 27 percent ignored or simply endured it, 7 percent retaliated physically against the bully or bullies, 5 percent sought protection from peers without telling them about the bullying, and 2 percent planned revenge. A study by Brown, Birch, and Kancherla (2005) surveyed 1,229 9- to 13-year-old students in seven states and found that when bullied, nearly half of the sample fought back, one-quarter reported the bullying to an adult, one-fifth did nothing, and fewer than one-tenth of the sample tried to talk to the bully about their behavior. Salmivalli et al. (1996a) summarized coping strategies by identifying three common responses of victims: counteraggression, helplessness, and nonchalance. Counteraggression involves reactions such as attacking the bully, defending oneself, provoking the bully, and calling for help; helplessness involves crying, doing nothing, and avoiding school; while nonchalance refers to reactions such as remaining calm, ignoring, and acting un-phased. Salmivalli and colleagues report that 70 percent of the victims in their study reacted with either helplessness or nonchalance, and counteraggression was more frequently used by boys than by girls. Salmivalli suggests,

*On the basis of these results, it is conceivable that when bullying starts, the bullies do not always expect the victim to submit…. In many cases, especially among boys, a powerless counterattack is expected. Maybe a completely helpless, for example, crying victim is not as rewarding prey to the bully than is a victim who can be provoked to ineffective, ridiculous counteraggression. This means not only an easy victory to the bully, but also a good "show" to the whole peer group.* (p. 108)

**Outcomes of reaction types**

The ways in which victims respond to their bullies is related to the course and outcome of the bullying episodes, as well as to different emotional outcomes for the victims. For example, Kochenderfer and Ladd (1997) found that boys who fought back against bullies
had stable rates of victimization over the kindergarten year, while those who enlisted a friend's assistance had reduced rates of victimization. In another study, Sharp (1996) found that students who reacted assertively or aggressively to bullying behavior experienced less stress when bullied than did students who reacted passively. Kochenderfer-Ladd and Skinner (2002) reported that fourth-grade boys who attempted to resolve peer arguments directly with bullies had lower levels of loneliness and fewer social problems, while boys who sought support from peers outside the conflict tended to be lonelier. Conversely, the authors found that seeking outside social support was an effective strategy for chronically victimized girls, with this strategy serving to reduce their risk of social problems. Avoidance strategies such as ignoring or trying not to think about the bullying were not particularly effective for boys or for girls, and trying to distance oneself from the problem was associated with greater anxiety for boys and associated with greater loneliness and more social problems for girls. The authors also found that for boys, externalizing strategies such as taking their frustrations out on others caused greater loneliness, while internalizing strategies such as blaming themselves were associated with loneliness, anxiety, and depression.

In a study of 120 elementary school victims of bullying, Mahady Wilton, Craig, and Pepler (2000) observed emotional regulation and display as the children interacted with peers during free play. The authors found that the two most common victim reactions during an episode of bullying were interest and joy, seen as dysfunctional responses in attention-craving children and tantamount to providing social reinforcement for bullies to continue with their aggressive behavior. The third most frequent response by victims was anger, which tended to be emotionally uncontrolled and to contain exaggerated displays of distress, leading to a prolonged bullying episode and to eventual defeat for the victim. The next most common response types were sadness and surprise, which the authors also considered maladaptive since they conformed to the bullies' expectations. Overall, Mahady Wilton et al. concluded that victims of bullying typically demonstrate deficits in emotional regulation and display, and lack the social resources to prevent the bullying from occurring and to generate protective responses from peers.
In another study of victims' emotional reactions, Berdondini (1999, as described in Cowie & Berdondini, 2002) assessed 35 children aged 8 to 11 years, including 13 bullies, 8 victims and 14 bystanders. The students were videotaped for one year as they participated in class-based group activities, after which they watched a video replay and were interviewed about their emotional responses to the actions seen on tape. The children’s facial expressions in the tapes were also analyzed and coded. The author determined that victims generally responded in one of two ways: passive victims responded with a "blank," unexpressive face, while aggressive victims displayed an over-reactive response. Bystanders reported being particularly disturbed by the expressionless response, which they perceived as "strange" or "not normal." Bullies expressed irritation when the individual they were victimizing did not show any overt emotion, and reported confusion over whether the victim was enjoying "the joke" or not. In some cases, Berdondini found that victims artificially emoted joy that was contradicted by their facial expressions and represented an attempt to deny their true feelings. This ambiguous response was sometimes seen as a challenge to bullies, and provoked continual attacks until a more satisfactory response, such as over-aggression or crying (a response usually abhorred by the peer group), was obtained.

Outcomes and Correlates of Victimization

Introduction

Scores of studies have been published that report on the impacts or outcomes of bully victimization, and virtually all such studies find deleterious effects for victims. The requisite challenge in these studies is determining whether the bully victimization itself is responsible for the observed negative outcomes, or whether victims of bullying are predisposed to such outcomes and would have experienced them in the absence of the bullying. Studies on bullying vary extensively in approach and methodological design. Longitudinal panel studies are one of the strongest methodological designs available to study the impacts of bullying; unfortunately such designs are time-consuming, expensive, and as such have not frequently been used. Generally, a panel design involves a baseline assessment of levels of bullying and victimization
behaviors, followed by repeated measure(s) at some monthly or yearly interval in the future. As such, the stability of behaviors over time can be assessed, and a temporal relationship between bullying/victimization and future outcomes can be established. Despite the difficulty in definitively establishing causal relationships without the use of a randomized control design (which is not feasible in bullying research because of ethical reasons), rigorous longitudinal panel designs with appropriate statistical controls may offer a valuable alternative for measuring bullying outcomes.

Another type of study commonly used to assess bullying and victimization outcomes is a recall study. Usually in the form of retrospective self-report surveys, recall studies involve adults responding to questions about their childhood experiences with bullying, as well as questions about their current behavioral and emotional functioning. Since their "outcomes" are known, adult responses regarding earlier bully victimization are used to mimic a pseudo-longitudinal study of bullying effects. These designs are limited by potential bias in respondents' recollections of previous experiences, for example, inaccurate remembering of event details, making erroneous attributions about other's behavior or the impacts of such behavior (for example, believing that negative scholastic outcomes were a result of the bullying, when in fact other variables may have been more influential), or a differential recall ability among respondents (for example, if anxious individuals retain memories of bullying in more vivid detail and recall a greater numbers of incidents than do non-anxious individuals, study results might suggest a spurious correlation between bullying and anxiety in adulthood). Memories of bullying in school may also be tainted by an individual's current circumstances, or incidents of bullying may have been forgotten over the years. Nevertheless, the handful of recall studies that exists provides an interesting perspective on bullying and victimization outcomes beyond the typical cross-sectional self-report surveys of children and youths.

Cross-sectional surveys are the predominant means by which researchers study bullying; as a result these studies make up the vast majority of the published literature. Cross-sectional studies typically involve a one-time questionnaire completed by students in a classroom, through which assessments of bullying and victimization are made along with
assessments of other variables related to behavior, emotions, health, or academic success. Given the nature of this type of design, the findings from such studies are only able to provide evidence of correlations among variables rather than convincing evidence of the causal impact of one variable on another. For example, many studies report a correlation between experiences of bully victimization and depression. It may well be that victimization causes children to feel rejected, lonely, and insecure, and these outcomes may in turn lead to depression. An alternative explanation for this finding is that children who are prone to depression tend to be withdrawn, submissive, and to display anxious behavior – all of which have also been demonstrated as risk factors for victimization. In other words, victimization may be a partial consequence of symptoms of depression. Correlational studies only allow for a relationship between variables to be established, they do not allow for a determination of whether victimization is responsible for any of the observed outcomes. Even so, correlational studies can provide valuable insights into the linkages between victimization and other variables, as long as readers are aware of the limited conclusions that can be drawn from such research.

Because the field of bullying crosscuts several research disciplines and because studies are conducted by authors ranging from classroom teachers to medical doctors to Ph.D.s in statistics, the level of rigor in methodological design of studies on bullying varies more dramatically than research in many other fields. The result of the diversity in quality across studies is that researchers seeking to synthesize these studies are faced with the challenge of combining rigorous, quantitative designs that involve several years of follow-up using large systematic samples and comparison groups with one-group post hoc designs undertaken in a single classroom of students. It is beyond the scope of this literature review to provide a comprehensive assessment of the methods and analysis strategies used in all bullying/victimization outcome studies, or to effectively weigh the results of rigorous studies more heavily than the results of less rigorous studies. Nonetheless, the fact that practically all studies, regardless of the methodology used, link bullying and victimization with negative outcomes strongly supports the contention that these occurrences are harmful to children.
**Longitudinal studies on victimization**

**Long-term longitudinal studies**

Although longitudinal follow-up studies of victims into adulthood are practically non-existent, those that exist suggest that some of the negative effects of bullying do not disappear with time. One such study tracked victims after their high school years and found that children who were bullied between sixth- and ninth-grade had much higher levels of depression and poorer self-esteem at age 23 compared with adults who were not bullied (Olweus et al., 1993d). The bullied individuals did not report higher levels of harassment or social rejection as adults than the individuals in the comparison group. In a Finnish longitudinal study, Sourander et al. (2006b) collected information on the problem behaviors of 8-year-old boys, and obtained follow-up information on criminal offenses from the National Police Register when subjects were 16- to 20-years-old. Results indicated that self-reports of peer victimization at age 8 were positively associated with future drunk driving offenses and violent crime.

**Short-term longitudinal studies**

i. Internalizing outcomes

Short-term longitudinal studies of school-aged subjects find that victims of bullying are more likely than non-victims to develop internalizing behaviors such as depression and anxiety (e.g., Fekkes et al., 2006a; Goodman et al., 2001; Hanish & Guerra, 2002; Kim, Leventhal, Koh, Hubbard, & Boyce, 2006; Kumpulainen et al., 1999; Marsh, Parada, Yeung, & Healey, 2001). For example, Kochenderfer & Ladd (1996a) examined the effects of peer victimization on school maladjustment in a sample of 200 kindergarten children. The authors found that victimization at the beginning of the academic year predicts loneliness and school avoidance at the end of the school year, with the duration of the victimization experiences related to the magnitude of subsequent school adjustment problems. Using the same sample of kindergarten students, Kochenderfer and Ladd (1996b) assessed the differential impacts of victimization type, finding that verbal victimization (both direct and indirect) was significantly related to
decreased liking of school, yet finding no links between school liking and physical victimization.

The link between victimization and internalizing behaviors has also been exhibited by older students. For example, Bond et al. (2001) surveyed 2,680 students in Victoria, Australia, twice in year 8 (at age 13 years) and again in year 9. After adjusting for sociodemographic factors, results indicated that recurrent victimization in year 8 was predictive of up to 30 percent of anxiety and depression for girls in year 9 (but not for boys). The authors also determined that newly reported victimization in year 9 was not significantly associated with year 8 reports of anxiety or depression, which suggests that victimization is predictive of emotional problems, but not the other way around. In other research, a one-year longitudinal study of 106 students in seventh- and eighth-grade found that students with concurrent perceptions of victimization had elevated levels of loneliness and low self-worth (Juvonen et al., 2000). Furthermore, a study by Sourander et al. (2006b) found that victimization at age 12 predicted ideations of deliberate self-harm at age 15, while Engstrom and Norring (2002) found that bully victimization nearly doubled the likelihood that a student in their last year of high school would be classified as high risk for an eating disorder.

The type of victimization experienced may be linked with the type of outcome found in subsequent assessments. In a one-year prospective study of the association of overt and relational victimization with social anxiety and social phobia in a sample of ninth-grade students, Storch et al. (2005) found that relational victimization predicted symptoms of social phobia (but not social anxiety). Overt victimization, on the other hand, was not a significant predictor of either outcome.

ii. Peer rejection

Peer rejection is another commonly identified outcome of bullying. Dill, Vernberg, Fonagy, Twemlow, and Gamm (2004) studied 296 third- through fifth-graders over two years, and found that victimization by peers at baseline predicted negative affect, peer rejection, and social skills deficits at follow-up. Peer rejection as a consequence of bullying was also reported by Hodges and Perry (1999) as well as by Khatri, Kupersmidt, and Patterson (2000).
iii. Delinquency outcomes

Deviant and delinquent behaviors are a concern for early victims of bullying (e.g., Khatri et al., 2000). For example, Kumpulainen and Rasanen (2000) found that victims of bullying at age 8 scored higher than non-involved children on future deviance (parent report), and on future bullying (teacher report). Hanish and Guerra (2002) followed a group of over 2,000 first-, second-, and fourth-graders over a period of two years, with a sample retention rate of 71 percent. After accounting for concurrent victimization and prior adjustment problems, the researchers concluded that victimization predicted later aggressive behavior, classroom attention problems, and delinquency. Similarly, Rusby, Forrester, Biglan, and Metzler (2005) followed 223 students in grades 5 – 7 into high school, reporting that students who experienced frequent peer harassment in middle school exhibited greater problem behaviors during their middle school and high school years. Students who were frequently verbally harassed in middle school displayed more antisocial behavior, deviant peer association, and alcohol use in high school, while those who experienced frequent physical harassment were more likely to display antisocial behavior, deviant peer association, and an increase in aggressive behavior.

iv. School outcomes

Longitudinal studies have linked bully victimization with negative academic outcomes (e.g., Schwartz, Gorman, Nakamoto, & Toblin, 2005). For example, Nansel, Haynie, and Simons-Morton (2003a) studied school adjustment among 930 youths surveyed at the beginning of sixth-grade, the end of sixth-grade, and at the end of seventh-grade. After controlling for baseline levels of school adjustment, youths who were victims during sixth-grade showed poorer school adjustment in seventh-grade and a more negative perception of school climate than did bullies and non-involved peers. Furthermore, bully victimization at school is related to a negative perception of peers and the view that schoolmates are hostile and less prosocial (e.g., Troop-Gordon & Ladd, 2005). In a study of 1,526 students in the fall and spring of the sixth-grade, Nishina, Juvonen, and Witkow (2005) found that peer
victimization in the fall predicted psychosocial maladjustment in the spring, and these outcomes predicted concurrent poor school functioning. Furthermore, psychosocial problems were found to mediate the association between peer victimization and school functioning. The authors suggest that victimized students were more depressed, anxious, and lonely than were students not subjected to frequent victimization, and these symptoms likely contributed to their disengagement from school.

v. Physical and psychosomatic symptoms

Finally, longitudinal studies have uncovered a relationship between bully victimization and physical symptoms. For example, a three-year study of 78 high school students in Australia (Rigby, 1999a) found that victimization at baseline was predictive of poorer physical health at follow-up for both males and females, using a measure that included symptoms such as colds, ear infections, headaches, rashes, sore throats, dizziness, stomachache, and wheezing. High levels of peer victimization at baseline also predicted poor mental health in girls, including poor coping, depression, anxiety, and psychosomatic symptoms. In another study, Nishina et al. (2005) found that peer victimization in the fall predicted physical symptoms in the spring, which in turn predicted concurrent poor school functioning. The authors suggest that victims of bullying may be at elevated risk for becoming physically ill, perhaps owing to an increase in stress hormones which suppress immune system functioning and increase the likelihood of colds or other illnesses. The authors also speculate that students may consider physical symptoms a more ego-preserving way to characterize their problems, and choose to describe themselves as "sick," rather than admitting to psychological adjustment difficulties.

In addition to longitudinal studies with children and youth, the possibility that long-term victimization may result in physiological brain changes has been proposed based on findings from animal studies. Research by Delville, Melloni, and Ferris (1998) of young male hamsters who were "bullied" found that such hamsters grew up to be more aggressive toward smaller hamsters than those who were not bullied. A more recent study by Berton et al. (2006) of repeatedly bullied mice (who were exposed to a different, large bully mouse every day for 10 days) found that the mice subsequently acted frightened even when caged with an unfamiliar
non-bully mouse. These reactions persisted for up to four weeks following exposure to the bullies. Furthermore, the victimized mice developed a long-lasting aversion to social contact, characterized physiologically via an altered mesolimbic dopamine system, specifically through the induced production of brain-derived neurotrophic factor (BDNF) in the reward circuit. Berton et al. contend that the bullying-induced BDNF may facilitate long-term neuronal changes by stimulating nerve cell growth, which causes the development of depression and results in social aversion. To test this theory the victimized mice were injected with a virus that knocks out BDNF production solely in the brain's reward circuit, after which the victim mice were no longer intimidated by the bully mice. Although generalizing the results of animal studies to humans may not be straightforward, these findings suggest the possibility that repeated victimization by bullies may result in physiological changes to brain chemistry.

**Recall studies on victimization**

A number of retrospective studies have been conducted in which adults are asked to recall their experiences of bully victimization during primary and secondary school. This body of research suggests a link between victimization and mental and emotional outcomes; for example, Araki (2002) concluded that "victims of childhood bullying were prone to have an inferiority complex in interpersonal situations and... they deal with interpersonal stressful events by means of non-engagement coping strategies, resulting in depression" (p. 55). A study by McCabe, Antony, Summerfeldt, Liss, and Swinson (2003) found a relationship between history of teasing or bullying experiences and current anxiety disorders. Specifically, a history of severe teasing experiences was more common among participants with social phobia (92 percent), compared with those suffering from obsessive compulsive disorder (50 percent) and panic disorder (35 percent). Furthermore, the reported severity of the bullying experience was related to age of onset of the anxiety disorder, with severe bullying linked to earlier anxiety problems. Studies also find a link between bully victimization and future levels of stress (e.g., Newman, Holden, & Delville, 2005), depression (e.g., Pirkola et al., 2005), loneliness (e.g., Schäfer et al., 2004; Tritt & Duncan, 1997), and low self-esteem (e.g., Schäfer et al., 2004). Matsui, Kakuyama, Tsuzuki, and Onglatco (1996), in a study of 134 Japanese male university students, found that
the influence of bullying depended on the victim's psychological adjustment prior to experiencing the victimization. The severity of victimization experiences was negatively correlated with current self-esteem and was positively correlated with depression only for respondents who previously showed low scores for self-esteem and high scores for depression.

Some authors suggest that victimization negatively impacts general life functioning. For example, a study by Fosse and Holen (2004) of 160 adult patients from a psychiatric outpatient clinic found that a history of bully victimization was linked to current reports of lower levels of education, receipt of social benefits, and poorer psychosocial adjustment. Similarly, a Finnish study of 68 participants in a training program for persons with long-term unemployment reported that a significantly greater proportion of program participants had experienced bully victimization compared with the rate reported by the general population (Varhama & Bjorkqvist, 2005).

With respect to physical ramifications, Gunstad et al. (2006) found that a history of being bullied and rejected by peers in school predicted adult obesity for men (but not for women), while a study by Striegel-Moore, Dohm, Pike, Wilfley, and Fairburn (2002) related bullying to binge eating disorder in women. Similarly, Shelton and Liljequist (2002) found that men with poor body image satisfaction (who considered themselves to be underweight or overweight) were more likely to have been victims of childhood bullying than were men who were more satisfied with their current body weight.

Negative psychosocial outcomes of victimization are a common finding in retrospective research. For example, a study of 170 college students found that retrospective reports of bully victimization were positively associated with current levels of shyness, and were inversely associated with the degree of trust reported in current friendships (Jantzer, Hoover, & Narloch, 2006). Similarly, a study of 414 college students found that reports of childhood teasing were related to poorer social self-esteem, lower levels of current intimacy and closeness, less comfort in trusting and depending on others, and increased concerns about being unloved or abandoned in relationships (Ledley et al., 2006). The positive association of victimization with attachment problems was stronger for students who were teased about
social, appearance, and performance issues compared with those teased about family background or academics. In a similar study, Matsuoka et al. (2006) reported that adult attachment was predicted by fewer childhood peer victimization experiences, while Schafer et al. (2004) found that while former victims of bullying desired to have emotionally close relationships, they had difficulty trusting and were somewhat uncomfortable getting close to others.

**Cross-sectional studies on victimization**

The results from cross-sectional research on bullying has linked victimization with a host of negative outcomes, including conduct problems such as weapon-carrying, fighting and substance abuse, school outcomes such as truancy and decreased academic performance, physical health symptoms such as headaches and stomach pains, and mental health outcomes such as anxiety depression, and suicide. As discussed previously, despite the existence of a positive correlation between victimization and a wide range of negative outcomes, the vast majority of these studies do not employ the requisite rigor in research design or analytic approach necessary to support a causal link between victimization and outcomes. Nevertheless, the sheer volume of studies that claim linkages between victimization and negative functioning strongly suggest that the experience of bullying is potentially harmful for victims.

**Mental health symptoms**

One of the most common findings in correlational studies is a link between victimization and internalizing behaviors. Victims are more likely than bullies or those not involved in bullying or victimization to display psychological and mental health symptoms, such as depression and anxiety (e.g., Baldry, 2004; Craig et al., 1998; Denny, Clark, Fleming, & Wall, 2004; Graham & Juvonen, 1998; Kaltiala-Heino, Rimpela, Rantanen, & Rimpela, 2000; Leadbeater, Boone, Sangster, & Mathieson, 2006; Nadeem & Graham, 2005; Nishina & Juvonen, 2005; Puura et al., 1998; Roland, 2002; Salmon, James, & Smith, 1998; Seals & Young, 2003; Storch, Zelman, & Sweeney, 2002). Such outcomes may represent a chain of negative
events, whereby repeated victimization leads to the development of increased anxiety and depression (Craig, 1998). A Finnish study found that boys who reported being frequently bullied were over five times more likely to have moderate to severe depression than were those who were not bullied, while frequently bullied girls were over three times more likely to be moderately or severely depressed than were those who were not bullied (Kaltiala-Heino et al., 1999). Furthermore, a British study found that peer victimization was a significant contributing factor for adolescent referral for psychiatric services, with 70 percent of the referrals resulting in a diagnosis of depression (Salmon, James, Cassidy, & Javoloyes, 2000; see also Haavet, Straand, Hjordahl, & Saugstad, 2005; Mills, Guerin, Lynch, Daly, & Fitzpatrick, 2004). In a related study, Lataster et al. (2006) found that non-clinical psychotic experiences were significantly associated with a history of school bullying, while Storch and Esposito (2003) found both overt and relational victimization to be positively associated with symptoms of posttraumatic stress (see also Mynard, Joseph, & Alexander, 2000).

Victims of bullying demonstrate increased levels of social withdrawal and isolation (Baldry, 2004; Schwartz, Chang, & Farver, 2001); unhappiness (Luster, Small, & Lower, 2002; Rigby & Slee, 1993; Williams, Chambers, Logan, & Robinson, 1996); bad temper (Due et al., 2005); irritability (Kumpulainen et al., 1998), low self-esteem (Graham & Juvonen, 1998; Mynard et al., 2000; Nadeem & Graham, 2005), self-blame (Graham & Juvonen, 1998); peer rejection and difficulty making friends (Crick & Bigbee, 1998; Kumpulainen et al., 1998; Nansel et al., 2001); social anxiety (La Greca & Harrison, 2005; Storch et al., 2002); lying (Salmon et al., 1998); nervousness (Due et al., 2005); helplessness (Due et al., 2005); indecisiveness (Kumpulainen et al., 1998), social avoidance (Storch et al., 2003; Storch et al., 2002); and loneliness (Forero, McLellan, Rissel, & Bauman, 1999; Graham & Juvonen, 1998; Hazler et al., 1992; Storch et al., 2003). Research also suggests that victims of bullying have difficulty initiating and maintaining successful relationships with romantic partners (Gilmartin, 1987).

In extreme cases, repeated rejection and victimization by peers have been linked to suicidal ideation and suicide attempts (e.g., Carney, 2000; Davies & Cunningham, 1999; Kim et
According to the Centers for Disease Control and Prevention (2008), suicide ranks as the third leading cause of death for American youth, after accidents and homicide. For example, an Australian study found that children who reported being bullied at least once a week were twice as likely as their non-bullied peers to "wish they were dead" or to disclose that they were repeatedly having thoughts about taking their own lives (Rigby, 1996). In another study, Cleary (2000) investigated the association between peer victimization and suicidal behaviors among 1,569 high school students in the state of New York. Cleary found that for both male and female respondents, all categories of suicidal behaviors were more common for students who reported having been victimized compared with those who had not been bullied.

**Psychosomatic and physical health symptoms**

Many victims of bullying develop psychosomatic and physiological symptoms, and some research indicates that victims are more likely to report experiencing lower levels of general health (Rigby, 1996). Examples of specific symptoms reported more frequently by victims than non-victims include stomach pains (Boey & Goh, 2001; Fekkes, Pijpers, & Verloove-Vanhorick, 2004; Williams et al., 1996), migraine and non-migraine headaches (Kumpulainen et al., 1998; Metsahonkala, Sillanpaa, & Tuomien, 1998); backache (Due et al., 2005; Natvig, Albreksten, & Qvarnstrom, 2001a); sleeping problems (Fekkes et al., 2004; Williams et al., 1996); poor appetite (Kumpulainen et al., 1998; Wolke, Woods, Bloomfield, & Karstadt, 2001a); bedwetting (Fekkes et al., 2004; Williams et al., 1996), dizziness (Due et al., 2005); tiredness (Fekkes et al., 2004), and repeated sore throats, colds, and coughs (Wolke et al., 2001a).

Engstrom, Hallqvist, Moller, and Laflamme (2005) studied 10- to 15-year-old Swedish adolescents who had been hospitalized or called back for a medical check-up because of a physical injury. The authors found bully victimization related to an increase in the short-term occurrence of unintentional injuries in childhood (excluding any injuries caused by bullies themselves or actions taken by the victim to avoid victimization). No long-term increase in the risk for unintentional injury was found. Furthermore, the risk of injury was much greater for
children who were infrequent victims of bullying than for those who experienced regular victimization (RR = 49.9 versus RR = 2.5; see also Laflamme, Engstrom, Moller, Alldahl, & Hallqvist, 2002).

**School outcomes**

For children who are victims of bullies, the school environment engenders fear and anxiety, which diminishes the potential for optimal learning. Victimization is linked to unhappiness and dissatisfaction with school, avoiding locations in the school/school grounds, increased absenteeism/truancy, difficulty concentrating, decreased academic performance, and school drop out (e.g., Berthold & Hoover, 2000; Brockenbrough, Cornell, & Loper, 2002; Forero et al., 1999; Glew, Fan, Katon, & Rivara, 2005; Nansel et al., 2001; Schwartz et al., 2001; Wolke et al., 2001b). For example, Miller, Verhoek-Miller, Ceminsky, and Nugent (2000) surveyed 117 college students and found bully victimization to be significantly negatively correlated with school satisfaction. An Australian study reported that 6 percent of boys and 9 percent of girls said they stayed at home to avoid being bullied at school (Rigby & Slee, 1999; see also DeVoe et al., 2005). Larger effects were found in a study of 5,489 children age 12-13 in schools across the UK, which indicated that a quarter of the boys and a third of the girls in this age group sometimes, often, or very often feared going to school because of bullying (Balding, 1996).

**Conduct problems**

In a nationally representative study of over 15,000 students in grades 6 through 10, those reporting weekly victimization were 60 percent more likely to carry a weapon to school for protection, 70 percent more likely to be involved in frequent fights, and 30 percent more likely to be injured than were students who did not report being bullied (Nansel, Overpeck, Haynie, Ruan, & Scheidt, 2003b). Victims have also shown elevated risks for underage drinking, smoking, and other substance abuse (e.g., Kuntsche & Gmel, 2004; Sullivan, Farrell, & Kliewer, 2006). As reported by the Safe on the Streets Research Team (1999), a study of 13,000 U.K. pupils under the age of 16 years found that nearly one quarter of child runaways
cited difficulties at school as forcing them to leave home, while 41 percent of runaways recounted past experiences of exclusion at school. Victims of bullying may also react to peer harassment with aggressive and violent behaviors (e.g., Hazler & Carney, 2000; Schwartz et al., 2001; Sullivan et al., 2006). Furthermore, a study by Hendriks and Bijleveld (2004) that compared juveniles who sexually offend against children with juveniles who sexually offend against peers or older victims found that juvenile sex offenders were more likely to have been bullied at school.

II. Bully-Victims/Provocative victims

Characteristics of Bully-Victims

A second, less common type of victim is the provocative victim, also known as the aggressive victim or "bully-victim." This type of victim is both harassed by aggressors and bullies others. While these children take on both the roles of bully and victim, they are primarily considered victims rather than bullies because their bullying behaviors are usually seen as maladaptive reactions to their experiences of victimization (e.g., Graham et al., 2006).7 Bully-victims are characterized by both anxious and aggressive reactions (Olweus, 1993c), are quick-tempered, and try to defend themselves or fight back when attacked (Limber, 2002). Hodges and Perry (1996) propose that bully-victims feel that injustices have been perpetrated against them, and they stand ready to blame, argue, lie, and seek revenge. Pellegrini (1998) describes bully-victims' reactions as follows:

Based on these hostile attributions, these children often react aggressively to others’ provocative social behavior. For example, an aggressive victim would typically attribute hostile and aggressive intent to a peer accidentally bumping into him…. Aggressive victims, unlike bullies, use aggression… in retaliatory circumstances, in response to what they perceive as a threat…. Unlike bullies, these children may not systematically choose weaker children as targets of their aggression. (p.168)

7 Some researchers categorize these children as "anxious bullies", e.g., see Miller, Beane, and Kraus (1998).
The bullying of provocative victims may involve many students or even the whole class, and the bully-victim’s aggressive responses serve to encourage further bullying by peers (Olweus, 1993c). Bully-victims have low self-esteem, attention problems, they may be hyperactive, and they tend to regularly irritate other students and cause tension and resentment (e.g., Kumpulainen, Rasanen, & Puura, 2000; Olweus, 1993c). These children have limited social skills and self-regulation skills; they tend to interrupt conversations, they lack patience, they have difficulty taking turns, and they display other behaviors consistent with attention deficit–hyperactivity disorder (e.g., Toblin, Schwartz, Gorman, & Abou-ezzedine, 2005). Bully-victims are among the least liked of their peer group, and often have few or no friends (e.g., Pellegrini, 2002; Perren & Alsaker, 2006; Perry et al., 1988; Unnever, 2005). Some research suggests that bully-victims experience more total bullying than do submissive victims (e.g., Dulmus, Sowers, & Theriot, 2006). Bully-victims are at particular risk for negative psychosocial outcomes (Haynie et al., 2001), as well as for engaging in various types of delinquent behavior. Brockenbrough (2002), in a survey of 10,909 seventh-, ninth-, and eleventh-grade students found that bully-victims were more likely than other students to have carried a weapon, consumed alcohol, and engaged in physical fighting at school.

A study by Unnever (2005) investigated differences in the behavior among 206 bullies, 514 victims, and 206 bully-victims in grades 6, 7, and 8 from six middle schools in Virginia. Unnever found that bully-victims were less proactively aggressive than were "pure" bullies, and were more reactively aggressive. Bullies and bully-victims did not differ in how often they harassed other students, but bully-victims were more likely to bully other students using physical means and were less likely to engage in verbal bullying. Furthermore, they were more likely to be the victims of physical bullying than were "pure" victims.

Prevalence of Bully-Victims

Although estimates vary, the prevalence of bully-victims cited in U.S. studies ranges from 1 percent to 30 percent. In their nationally-representative sample of sixth- to tenth-graders, Nansel and colleagues (2001) found that of the 17 percent of the sample who
indicated they were victims of bullying; about one-third could be classified as bully-victims. In another American study, Dulmus et al. (2006) studied nearly 200 children at rural elementary and middle schools, and found that 31 percent of the students were pure victims while 11.5 percent were more appropriately classified as bully-victims. Internationally, the proportions of bully-victims in student samples vary. A study by Menesini et al. (1997) surveyed 6,659 students in England and 1,376 students in Italy, reporting a prevalence rate of 1.2 percent in English middle schools and 7.0 percent in Italian middle schools. In primary schools, the rates were 4.5 percent (England), and 12.7 percent (Italy). In a Scottish study, Mellor (1990) found a lower rate, with only 3 percent of the sample classified as both bullies and victims. Conversely, an Irish study by O'Moore and Hillery (1989) found that 66 percent of bullies were also victims, and that 46 percent of victims were also bullies. Farrington (1993) suggests that the discrepancy in rates of bully-victims across studies is likely a result of differences in the stringency of measurement criteria for "bully" and "victim."

Bully-Victim Risk and Protective Factors

Certain family factors are thought to be related to the development of provocative victims, including parents who are low in monitoring and warmth, and inconsistent or abusive in their parenting roles (e.g., Bowers et al., 1994; Stevens, Bourdeaudhuij, & van Oost, 2002). For example, a longitudinal study involving mothers of 198 5-year-old boys found that at age 8-9 follow-up, those children who had experienced more punitive, hostile and abusive family environments, as well as inter-parental marital conflict during the child's preschool years, were more likely to be characterized as aggressive victims (Schwartz et al., 1997). These children differed from "pure" bullies in that pure bullies did not have home environments characterized by victimization from adults. Schwartz and colleagues propose this finding is evidence that:

...early experiences of victimization and harsh treatment by adults serve to dysregulate children emotionally, leading to later hyperactive anger and victimization by peers. Observation of violence among adults may dispose a child toward aggressive behavior, but only the experience of violence
disposes a child toward the combined problem of aggressive behavior and peer victimization. (p. 673)

This finding concerning abusive and conflictual home environments is not universal; Unnever (2005) did not find that bully-victims have more coercive home environments than do pure bullies, or that bully-victims are more frequently exposed to interparental conflict. Other research suggests that bullying by siblings may be an influential factor in the development of bully-victims. Duncan (1999b), using data from 375 middle school children in the rural mid-South, found that bullying by siblings at home was common for bully-victims, with 77 percent of bully-victims at school reporting involvement in both bullying and victimization with siblings at home. Additional research into the home environments of bully-victims is needed in order to more clearly differentiate them from pure bullies and pure victims.

Outcomes and Correlates of Bully-Victimization

Of the four roles in the bullying experience, bully-victims are the group that engenders the most concern and is at highest risk for both psychosocial problems and violence (e.g., Andershed, Kerr, & Stattin, 2001; Anderson et al., 2001; Duncan, 1999b; Haynie et al., 2001; Kaltiala-Heino et al., 1999; Kumpulainen & Rasanen, 2000; Nansel et al., 2001; Wolke et al., 2000). This consistent research finding suggest that bully-victims frequently display both the social-emotional problems of victimized children as well as the behavioral problems of bullies (e.g., Arseneault et al., 2006; Kumpulainen & Rasanen, 2000; Limber, 2002).

Similar to "pure" victims, bully-victims are at risk for increased levels of internalizing problems. In their study of sixth- to tenth-graders, Nansel and colleagues (2001) found that bully-victims reported more difficulty than either bullies or victims in developing friendships, suffered increased loneliness, and frequently encountered problems with their peers. Other studies find that bully-victims are considerably more depressed and anxious than are either pure bullies or bystanders (e.g., Fekkes et al., 2004; Kaltiala-Heino et al., 2000; see also Kaltiala-Heino et al., 1999), and are at increased risk for suicidal ideation (e.g., Kaltiala-Heino et al.,
A study of 420 children found that those involved in bully/victim problems were more likely than noninvolved children to have psychiatric disorders such as ADHD, oppositional/conduct disorder, and depression, although rates were not as high as for pure bullies (Kumpulainen et al., 2001). In addition, a study by Forero et al. (1999) of 3,918 Australian students aged 11 to 16 years found that bully-victims were significantly more likely to experience frequent and high scores for feeling low, nervous, and irritable. Furthermore, Kumpulainen and Rasanen (2000) reported that children who were characterized as bully-victims at the age of 8 years had greater negative psychiatric symptoms than other children, as well as higher levels of negative psychiatric symptoms at the age of 15.

Physical and psychosomatic health symptoms have been identified as correlates of bully-victimization (e.g., Kaltiala-Heino et al., 2000). For example, Fekkes et al. (2004) studied 2,766 elementary school children and found bully-victims to have an increased risk for abdominal pain, poor appetite, bed-wetting, and tiredness, while Wolke et al. (2001a) found that direct bully-victims were more likely to suffer symptoms such as repeated sore throats, coughs, and colds, and poor appetite. Similarly, a study of nearly 4,000 Australian students reported that bully-victims were significantly more likely than other students to experience symptoms such as headache, stomach ache, backache, and dizziness (Forero et al., 1999).

Negative associations of bully victimization with academic achievement have been observed, with some studies finding that bully-victims are at risk of lower academic achievement than are other students (e.g., Arsenault et al., 2006; Glew et al. 2005; Nansel et al., 2001). Research by Nansel et al. (2003a) also found that bully-victims have poorer school adjustment than students not involved in bullying, and were more likely to have a negative perception of the school climate (see also Arseneault et al., 2006; Forero et al., 1999).

Externalizing problems are a serious concern for bully-victims, with such children often showing increased levels of aggression, conduct disorder, and disobedience than other students (e.g., Craig, 1998; Kumpulainen & Rasanen, 2000; Wolke et al., 2000). For example, Kim et al. (2006) studied 1,655 seventh- and eighth-grade Korean students, and reported that students identified as bully-victims at baseline had increased levels of aggression and
externalizing problems after 10 months. A study of 5,813 elementary school-aged children found that bully-victims scored higher than other children for externalizing problems such as fighting, disobedience, lying, irritability, temper tantrums, and stealing, as well as hyperactivity (Kumpulainen et al., 1998).

Research links bully-victimization with various delinquency outcomes, including substance abuse and drinking alcohol (Brockenbrough et al., 2002; Kaltiala-Heino et al., 2000; Nansel et al. 2001), smoking (Forero et al., 1999), and carrying weapons (Andershed et al., 2001; Brockenbrough et al., 2002). A study by Huizinga and Elliott (2002) found that male bully-victims were twice as likely as other males to be involved in future interpersonal violence, with over half of the bully-victims in the sample reporting involvement in minor or serious assaults. Bully-victims' aggressive reactions to incidents of peer harassment may also put others at risk. The most extreme consequence of bullying is violence to self or others, and bully-victims are those children most likely to commit suicide and homicide (e.g., A. Pellegrini, cited in Viadero, 2003). Columbine shooters Dylan Klebold and Eric Harris were labeled as bully-victims by an investigative panel that sought to draw lessons from the 1999 school shootings (Viadero, 2003). A report by the U.S. Secret Service supports the contention that school shooters are likely to be bully-victims. Vossekuil, Fein, Reddy, Borum, and Modzeleski (2002) examined 37 incidents of targeted school shootings and attacks involving 41 attackers that occurred between December, 1974 and May, 2000, amassing all available information about each incident and conducting in-depth interviews with 10 of the shooters. The authors found that "almost three-quarters of the attackers felt persecuted, bullied, threatened, attacked or injured by others prior to the incident" (p. 21). In several of the cases studied, the attackers experienced bullying that was severe, long-standing, torturous, and in more than half of the cases revenge played a significant role in the attacker's decisions to resort to violence. For example, in one case witnesses indicated that "... nearly every child in the school had at some point thrown the attacker against a locker, tripped him in the hall, held his head under water in the pool or thrown things at him" (p. 21). The Secret Service report further stated "these attackers told of behaviors that, if they occurred in the
III. Bullies (Perpetrators)

Aggressive versus Passive Bullies

**Aggressive bullies**

While victims are described as being submissive or provocative, bullies are characterized as aggressive or passive (Carney & Merrell, 2001; Coy, 2001). Aggressive bullies are the most well-known, and tend to be impulsive, belligerent, coercive, confident, and tough, with a lower tolerance for frustration than other children and a stronger inclination towards violence (e.g., Bosworth, Espelage & Simon, 1999; Coy, 2001; Olweus, 1993c; Perren & Alsaker, 2006). These children are quick to anger and resort to the use of force, and may view aggression as the only way to preserve their tough self-image and resolve conflicts (Hazler et al., 1997). Bullies of this type usually behave aggressively to both peers and adults, have dominant personalities, enjoy being in control, have difficulty conforming to rules and obeying authority, are less cooperative than other children, and lack empathy towards their victims (e.g., Coolidge, DenBoer, & Segal, 2004; Perren & Alsaker, 2006). Olweus (1978) contends that aggressive male bullies tend to be physically stronger than average, and are likely to be stronger than their victims. Some research suggests aggressive bullies have decreased levels of salivary cortisol, a chemical that is usually released in response to fear, for example, the fear that might arise with respect to potential consequences for negative behavior (e.g., McBurnett, Lahey, Rathouz, & Loeber, 2000).

**Passive bullies**

Aggressive bullies are often surrounded by a group of two or three "passive" bullies who support them in their behavior (Cairns, Cairns, Neckerman, Gest, & Gariépy, 1988; Olweus et al., 1993c). As a group, the aggressive, more powerful bully instigates the bullying...
episode while the passive bullies (also known as "henchmen") reinforce the behavior and begin to actively participate once the bullying is underway (Coy, 2001). Passive bullies rarely initiate aggressive acts themselves, but perpetuate the continuation of bullying behavior by providing the aggressive bully with approval and admiration. In turn, passive bullies are inspired to carry out their own bullying behaviors after seeing an aggressive bully receive respect and rewards from other passive bullies. Ross (2002) characterizes passive bullies as being intensely loyal to aggressive bullies, "If, for example, adults intervene in bullying that is initiated by the aggressive bully and impose sanctions, the [passive] bullies are often blamed and then accept the punishment without any attempt to implicate the aggressive bully" (p. 114).

Characteristics of Bullies

In general, bullying behavior is associated with externalizing problems such as conduct disorder, oppositional defiant disorder, and ADHD (e.g., Coolidge et al., 2004; Haynie et al., 2001). Bullies often have difficulty controlling their impulses and emotions, along with significant problems in managing anger (e.g., Haynie et al., 2001; Pellegrini et al., 1999). Some research suggests that bullies are sensation-seeking by nature, as a result of being cortically under-aroused. For example, in a study of 242 secondary school students with a mean age of 13.5 years, Woods and White (2005) found that direct bullies and non-involved children had the lowest average levels of arousal, and that direct bullies were less frequently characterized by clinically high arousal levels. The authors suggest that direct bullies may benefit from a low level of arousal, which enables them to control their behavior and manipulate others strategically; waiting to initiate aggressive confrontations until presented with an opportunity for successful interaction. Contrary to expectations, Woods and White found that relational bullies had higher clinical levels of arousal than did victims and non-involved children. The authors surmise that since relational bullying does not involve the one-on-one confrontation that is characteristic of direct bullying, engaging in such behavior does not produce the same increase in arousal. Since a direct bully seeks a boost in arousal as a result of bullying, children
who engage in direct bullying may be intrinsically different from those who engage in relational bullying.

Bullies do not usually limit their harassment to a single victim. A British study found that more than two thirds of perpetrators bullied multiple victims at any given time, while over three quarters bullied different victims at different times (Stephenson & Smith, 1989, as cited by Farrington, 1993). Accordingly, victims tended to be harassed by more than one bully in the same episode, and were victimized by different perpetrators at different times. Chan (2006) describes results from his 2002 dissertation, in which he examined systemic patterns of school-wide bullying and victimization. Chan identified a pattern of serial bullying in his data, in other words, single bullies who harassed two or more victims, often targeting students in different classrooms or grades. While serial bullies were small in number, Chan found they were responsible for a large portion of the observed episodes of bullying, and surmised that intervention efforts with these students would have a large impact in reducing bullying school wide.

Gender

Many studies find that boys are more likely than girls to bully others (e.g., Baldry, 2003; Craig, Peters, & Konarski, 1998; Duncan, 1999b; Nansel et al., 2001; Natvig, Albrektsten, & Qvarnstrom, 2001b; Olweus, 1993c; Pellegrini & Bartini, 2000a; Whitney & Smith, 1993). For example, a study involving students in grades 4 though 8 found that 23 percent of boys versus 8 percent of girls had bullied others more than once or twice a term (Charach et al., 1995). Likewise, the nationally representative study by Nansel et al. (2001) found that 26 percent of boys engaged in moderate to frequent bullying, compared with 14 percent of girls. Alternative methods of measurement have confirmed a gender difference; a study using naturalistic observations of bullying on the school playground observed boys bullying at a rate of 5.2 episodes per hour compared with an hourly rate of 2.7 episodes for girls (Craig & Pepler, 1997).

Gender differences in rates of bullying are not universally found; research by Atlas and Pepler (1998) reported no differences in bullying for girls versus boys, while Bauer et al. (2006)
found higher rates of bullying among girls (female to male ratio of 1.5:1). Bernstein and Watson (1997) reviewed the literature and concluded that gender differences are likely a result of inconsistent definitions of bullying used across studies, such as whether respondents are provided with examples of both physical and verbal aggression. For example, a study of 65 preschoolers found that boys were significantly more likely to be overtly aggressive than were girls; however, girls were significantly more likely to be relationally aggressive (Crick et al., 1997). Other studies also find that girls are more likely to be relationally aggressive (e.g., see Crick & Nelson, 2002). If respondents are unaware (or are unwilling to admit) that some relationally aggressive behaviors constitute bullying, prevalence rates may be biased towards underreporting for girls.

Mixed results have been found regarding the gender of bullies in relation to their victims. Some studies report a predominance of within-gender bullying, i.e., that boys target boys more often than they target girls, while girls target girls more often than they target boys (e.g., Pellegrini & Long, 2002). Other studies find that while girls bully/are bullied by both boys and girls, boys typically bully/are bullied only by other boys (e.g., Olweus, 1993c; Whitney & Smith, 1993). For example Boulton (1996), in a sample of 192 8- to 10-year-old children found that boys were significantly more likely to select boys rather than girls as victims, but girls were not significantly more likely to target girls over boys. Similarly, Craig and Pepler (1997) found that 86 – 89 percent of the victims of male bullies were male. For female bullies substantially lower within-gender dyads were found, ranging from 48 to 52 percent.

Marked gender differences are reported with respect to specific types of bullying. Boys are more likely to report physically bullying other students (e.g., Barone, 1997; Crick & Nelson, 2002; Nansel et al., 2001; Tomada & Schneider, 1997; Whitney & Smith, 1993). Girls are more likely to engage in behaviors that are covert and manipulative, such as rumor-spreading, exclusion, and manipulation of friendships (for example, depriving a girl of her "best friend") (e.g., Crick et al., 2006; Crick & Nelson, 2002; Lagerspetz et al., 1988; Nansel et al., 2001; Olweus, 1993c; Underwood, 2003; Whitney & Smith, 1993). Girls may use relational aggression to gain or preserve a favorable social position among their peers (Merten, 1997),
and research indicates that this type of aggression occurs frequently – even among close friends (Crick & Nelson, 2002). Highlighting the gender difference, Crick and Nelson (2002) determined that failing to include measures of relational aggression among friends in addition to measures of physical aggression would have missed the identification of 71 percent of female victims and 21 percent of male victims in their study.

Some research refutes the finding that females engage in more relational aggression than do males, e.g., Rys and Bear (1997) studied 266 third- and sixth-graders in the U.S. and found more overt aggression among boys, but no gender differences in relational aggression (see also Dhami, Hoglund, Leadbeater, & Boone, 2005; Espelage, Holt, & Henkel, 2003; Storch, Masia-Warner, Crisp, & Klein, 2005; Tomada & Schneider, 1997). Similarly, a study by Loukas, Paulos, and Robinson (2005) found that boys reported higher mean levels of overt aggression than did girls, but there were no significant differences between boys and girls for levels of social aggression. In a finer level of analysis, the researchers found that more boys than girls were classified as overtly aggressive (13.3 percent of boys versus 5.3 percent of girls) and combined socially/overtly aggressive (8.9 percent of boys versus 3.3 percent of girls), but more girls than boys were exclusively socially aggressive (11.3 percent of girls versus 6.6 percent of boys). A recent meta-analysis of gender differences in aggression suggests that differences found in relational aggression are influenced by the method of measurement (Archer, 2004). Specifically, Archer reported that indirect aggression was more prevalent in females than males when using direct observational methods (d=.74), followed by peer ratings (d=.19), and teacher reports (d=.13).

Self-esteem

Controversial in bullying research is whether children who bully have low self-esteem and suffer feelings of inadequacy; the idea being that bullies pick on victims to compensate for their low level of self-regard. Some researchers find that bullies do have low self-regard (e.g., Connolly & O'Moore, 2003; Ma, 2001; O'Moore & Kirkham, 2001; Rigby & Cox, 1996), while others report that bullies have average or high self-regard in comparison with their classmates (e.g., Graham et al., 2006; Johnson & Lewis, 1999; Natvig et al., 2001b; Olweus, 1993c, Pearce
& Thompson, 1998, Salmivalli et al., 1999; Vaillancourt, Hymel, & McDougall, 2003). For example, O'Moore and Kirkham (2001) conducted a nationwide study in Ireland using a sample of 8,249 children aged 8-18 years, and found that pure bullies had higher self-esteem than bully-victims, but had significantly lower self-esteem than children not involved in bullying. Similarly, in a study of 281 14- and 15-year-olds, Salmivalli (1998) found that bullies typically had high social and physical self-concept. Slee and Rigby (1993b) reported opposing findings in a study of 214 Australian boys. The authors found that victims, but not bullies, showed significantly lower self-esteem than did non-involved children. Smith (2004) suggests that the disparate findings with respect to self-esteem may be due in part to differences in study methodologies or samples.

The relationship of self-esteem with bullying may be somewhat more complex. O'Moore (2000) proposes that lower anxiety levels among bullies disguises their feelings of inadequacy and explain why bullies are perceived by others as confident (see also O'Moore & Kirkham, 2001). Research by Baumeister, Smart, and Boden (1996) proposes that violent, aggressive individuals may think highly of themselves but are also very sensitive to any criticism; in other words, their inflated beliefs of superiority (rather than legitimately high self-esteem), might invite threats from others, leading these individuals to engage in defensive aggression.

**Popularity**

Bullies are often as popular as other students (e.g., Cairns et al., 1988; Olweus, 1978; Slee & Rigby, 1993a; Vaillancourt et al., 2003). For example, Nansel et al. (2001) reported that students who bullied their peers had an easier time making friends than did non-bullying students. Juvonen et al. (2003) found that bullies were rated by their classmates as "cool," which may suggest a level of admiration from peers who yearn for self-confidence and a position of high social status in the school hierarchy. Accordingly, some studies find that bullies are perceived by their peers as powerful and possessing of attractive leadership qualities (e.g., Perren & Alsaker, 2006; Vaillancourt et al., 2003). Despite their ratings of high social status, Juvonen et al. (2003) found that students tend to avoid the company of bullies, which
suggests that their status may be more a result of intimidation rather than genuine affection. As such, bullies tend to have reciprocal friendships mostly with other bullies (e.g., Pellegrini et al., 1999; Perren & Alsaker, 2006).

This finding on popularity differs from research that observes chronically aggressive children to be less popular with their peers (e.g., Foster, deLawyer, & Guevremont, 1986; Kaufmann, 1985). Hoover and Hazler (1991) suggest that this difference may be because bullies do not behave aggressively towards all of their peers; rather, they selectively pick out targets. A bully's popularity may not be lasting, however, as some research observes that as bullies progress in school their popularity decreases until, by about age 15 or 16, they are considerably less popular than average (Olweus, 1997). Despite this decrease, bullies never attain the low level of popularity held by many chronic victims (Olweus, 1994).

**Social skills**

A commonly held belief is that bullies have a poor understanding of others; that they are socially inadequate and turn to aggression because they are unaware of how to interact properly with their peers (i.e., the social skills deficit model of aggression, see Crick & Dodge, 1994). Much research suggests that, in fact, bullies have high social intelligence skills which enable them to manipulate others and use covert bullying techniques to avoid detection (e.g., Kaukiainen et al., 1999; Sutton, Smith, & Swettenham, 1999a). As described by Kaukiainen and colleagues (1999):

> Efficient use of indirect aggression presupposes understanding of human relations and skills, which are applied in social settings. To use indirect aggression, the individual must be able to put his/her intentions to harm another person in a favorable light. At the same time, s(he) has to interpret the reactions of others and accommodate his/her behavior for the social manipulation not to backfire. These are all demanding skills....” (p. 88)

Sutton, Smith and Swettenham (1999b) found that when compared with passive bullies, victims, and peers who choose to defend victims, aggressive bullies demonstrate superior performance on a set of tasks that distinguishes between cognitive skills and
emotional understanding, Sutton and colleagues contend that bullies with superior "theory of mind" skills are at an advantage in their aggressive interactions with others. Furthermore, research by Perry et al. (1990) suggests that bullies can readily differentiate between children who are good for victimization and those who are not, and that bullies expect archetypical victims to have outcomes such as outward signs of suffering and a lower likelihood of retaliation.

**Expert agreement on bully characteristics**

As mentioned previously in the discussion of victim characteristics, Hazler et al. (1997) conferred with a group of 14 bullying experts on the importance of each of 70 characteristics at identifying bullies and victims. Table 3 presents results for the 19 key characteristics of bullies that were supported by at least 10 out of the 14 experts (71 percent).

**Table 3 Characteristics of Bullies as Rated by Experts**

1. Control others through verbal threats and physical actions
2. Quicker to anger and use force than others
3. Little empathy for the problems of the victim
4. Have often been exposed to models of aggressive behaviors
5. Chronically repeat aggressive behaviors
6. Inappropriately perceive hostile intent in the actions of others
7. Angry, revengeful
8. Parents are poor role models for getting along with others
9. Likely to have contact with aggressive groups
10. Parents are poor role models for constructively solving problems
11. See aggression as the only way to preserve their self-image
12. Inconsistent discipline at home
13. Perceived physical image is important for maintaining a feeling of power and control
14. Focus on angry thoughts
15. Many more family problems than usual
16. Parents often do not know the child's whereabouts
17. Suffer physical and emotional abuse at home
18. Create resentment and frustration in their peer group
19. Exhibit obsessive or rigid actions

*Adapted from Hazler et al. (1997), p. 9
 Persistence of Bullying Behaviors

Bullying is not a phase that all children grow out of, and for some bullies the behaviors persist over time (e.g., Camodeca et al., 2002; Kumpulainen et al., 1999; Pellegrini & Bartini, 2000a; Schafer, Korn, Brodbeck, Wolke, & Schulz, 2005). Olweus (1978) found that two-thirds of bullies maintained their bully status when assessed the following year, while Bjorkvist, Ekman, and Lagerspetz (1982) reported that 95 percent of bullies were rated by their peers as having been bullies in the previous year. Similarly, Sourander et al.'s (2000) study in Finland found that those children characterized as bullies at age 8 were also characterized as bullies at age 16 follow-up. Farrington (1993), in a longitudinal study of boys from ages 14 to 32, determined that bullying is persistent in individuals and may have an element of intra-generational continuity. Farrington discovered that boys identified as bullies at age 14 were also prone to be bullies at ages 18 and 32. Furthermore, at age 32 follow-up those boys characterized as bullies at age 14 were likely to have children who were bullies as well, even after accounting for the general tendency of antisocial fathers to bear antisocial children.

Other research finds lower rates of bullies in samples of older adolescents than in samples of younger children. In their study of over 15,000 students, Nansel and colleagues (2001) found that older youths were less likely to report bullying their peers than were youths in the younger age category. It may be that bullying declines with age because over the years youths gradually acquire better social skills, and because as children age the proportion of older, stronger bullies in their peer groups decreases (Smith et al., 1999). The latter suggestion is supported by research which claims that older children may target their harassment on younger, weaker children (e.g., Olweus, 1993c).

 Bully Risk and Protective Factors

There is no single cause of bullying behavior, and no predominant factor that predicts which children become bullies. Rather, the interaction of individual, familial, peer, school, and community factors serve to increase the risk that a child will become a bully (Allen, 2005; Espelage, Bosworth, & Simon, 2000; Olweus, 1993c).
Individual factors

Difficulty with self-control is a significant risk factor for bullying, as are difficulties with regulating attention and activity (e.g., Hay et al., 2004). For example, in a study of 1,315 middle school students, children who had previously taken medication for ADHD were found to have low levels of self-control and to be at increased risk for bullying (Unnever & Cornell, 2003). The authors also suggest that the link between physical strength and bullying may be mediated by level of self-control. Students with low self-control and a high perception of strength were found most likely to bully others, while students with high self-control and high strength were not more likely to engage in bullying. Related to strength is physical size, with larger boys being more likely than smaller boys to engage in peer victimization. For example, Griffiths et al. (2006) found that obese English boys were 1.66 times more likely to be overt bullies than were average weight boys (see also Janssen, Craig, Boyce, & Pickett, 2004, for similar findings in a Canadian study).

Other individual risk factors for bullying include an anxious-avoidant attachment pattern (e.g., Troy & Sroufe, 1987), display of externalizing behaviors (e.g., Bollmer et al., 2005), and anger. For example, Espelage and Swearer (2003) found that middle school students who were rated as angriest in the fall semester were found to engage in higher rates of bullying six months later (Espelage et al., 2001; see also Bosworth et al., 1999 and Coolidge et al., 2004). Similarly, Baldry and Farrington (2005) found that children are more likely to engage in bullying behavior if they cope with problems by getting angry and nervous. Conversely, the authors found that using problem solving coping skills is negatively associated with bullying.

Familial factors

Bullies often come from hostile family environments with authoritarian parents who use harsh and inconsistent methods of discipline, or overly permissive parents who provide minimal supervision of children (e.g., Ahmed & Braithwaite, 2004; Allen, 2005; Baldry & Farrington, 2005; Bowers et al., 1994; Craig et al., 1998; Myron-Wilson, 1999; Olweus, 1994; Schwartz et al., 1997). For example, Berthold and Hoover (2000) studied a sample of 591 students in grades 4 to 6 in a mid-sized Midwestern city and found that bullies were more
likely than other students to spend unsupervised time at home. Strassberg, Dodge, Pettit and Bates (1994) found that children who were disciplined by their parents with spanking showed higher levels of aggression toward peers than did non-spanked children. As concluded in an early study by Olweus, "too little love and care and too much "freedom" in childhood are conditions that contribute strongly to the development of an aggressive reaction pattern" (Olweus, 1994, p. 1182).

The families of children who bully may have poor problem-solving abilities, and parents may lack warmth, emotional support, and involvement in their children's lives (Allen, 2005; Baldry & Farrington, 2005; Batsche and Knoff, 1994; Myron-Wilson, 1999; Olweus, 1994; Rigby, 1994). A study by Curtner-Smith (2000) of 54 mother-son dyads found that mothers of bullies lacked social skills when interacting with other adults as well as with their own children, and that these poor social skills were modeled by their sons. Berdondini and Smith (1996) studied 60 Italian students aged 8-11 years and found a link between bullying and a high level of father absence, while Flouri and Buchanan (2003) found that bullying behavior was predicted by low father involvement and by low mother involvement. Farrington (1993), in a 24-year-longitudinal study, identified the most important predictors of bullying as familial, including physical neglect at age 8, having parents convicted of a crime at age 10, and having a father who did not participate in the boy's leisure activities at age 12.

A number of studies find that families of bullies have a high level of interparental conflict (e.g., Mohr, 2006). An assessment of the effects of interparental violence on children's involvement in bullying was conducted on a sample of 1,059 Italian elementary and middle school students (Baldry, 2003). Physical violence between parents was significantly related to direct bullying in children, with bullies being 1.8 times more likely to have been exposed to domestic violence than not exposed. The impact of mothers threatening fathers was also a significant predictor of bullying, with children exposed to interparental threats being more likely to be involved in indirect bullying in school. Some research also points to a link between the experience of child maltreatment (physical and sexual abuse) and bullying behavior, suggesting that the bully at school may be a victim at home. For example, Shields and Cicchetti

---

9 Low academic achievement at age 11 was also identified as an important predictor of bullying.
studied 169 maltreated and 98 non-maltreated boys and girls attending a summer day camp, and found bullying to be more common among maltreated children than among non-maltreated children. However, the direct impact of child maltreatment is unclear. Schwartz et al. (1997) conducted a longitudinal assessment of 198 5-year-old boys and found that exposure to adult aggression and conflict, but not direct victimization by adults, significantly predicted bullying behaviors.

As well as possible victimization in the home by parents, Duncan (1999b) reported that victimization by siblings is common. In a study of 375 rural middle school students, Duncan found that 41 percent of respondents bullied their siblings, while 30 percent were frequently victimized by their siblings. The results were closely linked to rates of bullying at school; 56 percent of school bullies reported that they also bullied their siblings at home. Some research suggests that children with a higher number of siblings are more likely to bully others (e.g., Eslea & Smith, 2000; Ma, 2001).

Protective factors related to the family context include supportive and authoritative parents (e.g., see Baldry & Farrington, 2005; Mohr, 2006). In a study of 1,147 British adolescents aged 14-18 years, Flouri and Buchanan (2003) found that both father involvement and mother involvement in their child's life were significantly associated with lower rates of bullying behavior, with neither parent appearing to have a more influential impact than the other. However, father involvement was more important in preventing bullying when mother involvement was low. After controlling for racial/ethnic background, gender, and age, Zimmerman et al. (2005) found that parental cognitive stimulation and emotional support for their 4-year-old children both served as protective factors against bullying when children were between the ages of 6 and 11 years.

**Interpersonal factors**

Similar to the findings for victims, friendships play an important role in bullying. Bollmer and colleagues (2005) found that children who displayed externalizing behavior problems were more likely to engage in bullying behavior if they had a low quality best friendship compared with children who displayed externalizing behaviors but had a high
quality best friendship. The authors theorize that high quality friendships – characterized by closeness, security, and helping – might help children practice social skills, increase their levels of empathy, and improve their ability to perceive others' feelings. Some research indicates that bullies start dating earlier than other adolescents, participate in more types of dating activities, are more likely to have a current girlfriend (males), and spend more time outside of school with other-sex friends and their boyfriends or girlfriends (Connolly, Pepler, Craig, & Taradash, 2000). However, Connolly et al. reported that bullies perceived relationships with their boyfriends or girlfriends to be less intimate, less affectionate, less equitable, and less durable than did comparison youths. Bullies were also more likely to report acts of physical aggression and social aggression against their partners than were comparison youths.

**School and community factors**

School- and community-related risk factors for bullying include a lack of adequate adult supervision, together with cultures in which teachers, other adults, and students have indifferent or tolerant attitudes towards bullying (Natvig et al., 2001b; Olweus et al., 1999). With respect to adult supervision, a study by Gage, Overpeck, Nansel and Kogan (2005) reports a relationship between bullying and the number of evenings spent out with friends. The authors obtained data from a U.S. nationally representative sample of 14,818 youths in grades 6-10, and found that 20 percent of youths spent five or more evenings out with friends each week. After controlling for variables such as age, racial/ethnic background, and parental education, these youths were significantly more likely to have engaged in a range of delinquent activities including fighting, carrying a weapon, smoking, and drinking alcohol. Both girls and boys were also significantly more likely to engage in bullying at school.

Resnick et al. (1998) maintain that the most influential predictor of adolescent health and well-being is perceived school connectedness, i.e., that students feel they are fairly treated, that they have friends they are close to, and that they feel devoted to their school. For example, Natvig et al. (2001b) found an increased risk of bullying associated with an increased level of school alienation, defined as the degree to which students feel their school work is meaningless and unchallenging. In another study, Graham et al. (2006) report that bullying is
more prevalent among youth who perceive school rules to be unfair. Schools that do not provide sufficient negative consequences for bullies perpetuate the problem and compound the risk factors that children bring to the classroom (Ma, 2001; Reinke & Herman, 2002), as do adults who believe that bullying is a normal (and therefore acceptable) part of growing up, and bystanders who admire bullies' strength or watch without intervening.

Outcomes and Correlates of Bullying

**Delinquency, crime, and conduct problems**

Longitudinal research suggests that bullying behavior in childhood is predictive of future aggression and conduct problems, with bullying serving as a "gateway" to antisocial and delinquent behaviors (e.g., Ferrell-Smith, 2003; Khatri et al., 2000; Kim et al., 2006; Kumpulainen & Rasanen, 2000). For example, a Norwegian study assessing age-24 outcomes in a sample of males reported that 60 percent of students characterized as bullies in grades 6 to 9 had at least one criminal conviction by the age of 24 (Olweus, 1993d). Even more striking was the finding that 35 to 40 percent of the former bullies had three or more convictions by age 24, compared with 10 percent of those who were neither victims nor bullies. Olweus's findings have been widely cited, as they suggest the prognosis that bullies are three to four times as likely as their non-bully and non-victim peers to have multiple convictions by their early 20s. Similarly, a study of 870 third-graders in the Midwest determined that bullies at age 8 had a 25 percent chance of acquiring a criminal record by age 30, compared with a rate of 5 percent for non-bullies (Eron & Huessman, 1990). The former bullies were also more likely to have been convicted of serious crimes and total overall crimes, to have dropped out of high school, and to have abused their wives and children. In another study, Sourander et al. (2006b) collected self-report data on bullying at age 8, and found that 8 to 12 years later, when subjects were 16- to 20-years-old, prior bullying predicted drug offenses, violent offenses, property offenses, traffic offenses, and drunk driving. In addition, a retrospective study of 10,000 incarcerated offenders from 85 correctional institutions found that a history of early bullying was a significant predictor of future gang membership (Holmes & Brandenburg-Ayres, 1998).
Bullying is frequently associated with externalizing behavior problems (e.g., Laukkanen, Shemeikka, Notkola, Koivumaa-Honkanen & Nissinen, 2002; Wolke et al., 2000), general aggressive behavior (e.g., Craig, 1998; Craig et al., 1998) and violence (Killias & Rabasa, 1997; Nansel et al., 2001). Cross-sectional studies have found bullying to be related to other antisocial behaviors such as drinking alcohol (Haynie et al., 2001; Nansel et al., 2001), drug use (Kaltiala-Heino et al., 2000), cigarette use (Forero et al., 1999; Haynie et al., 2001), gang membership (Knox, 1997); theft (Haynie et al., 2001), vandalism (Farrington, 1993; Haynie et al., 2001), property crime (Craig et al., 1998), and domestic violence (Farrington, 1993). For example, a study by Andershed et al. (2001) of nearly 3,000 14- and 15-year-old youths in Sweden found that youths who reported bullying others in school were substantially more likely to report committing violent behavior, carrying weapons, and being the victims of violence than were youths who were not bullies at school.

An outcome of particular concern in light of the potential for school violence is weapon carrying, a behavior reported more frequently among bullies than among non-bullies. Nansel et al. (2003b) found that bullies were over five times more likely to carry weapons to school than were children not involved in bullying, while the most persistent bullies (those who bullied at least once a week and who also bullied away from school grounds) were seven times more likely to report carrying a weapon to school. A recent study of over 6,000 students in grades 5 through 7 in rural South Carolina examined the relationship between bullying and gun ownership (Cunningham, Henggeler, Limber, Melton, & Nation 2000). Gun owners categorized as high risk (i.e., those who admitted that they owned guns to gain respect or to frighten others), reported higher rates of bullying than did those who did not own guns and those who were low-risk gun owners (i.e., those who owned guns to feel safe or to use in hunting or target-shooting).

The type of bullying exhibited may be related to the development of long-term delinquency outcomes. Huizinga and Elliott (2002), in a study of over 1,500 youths participating in the prospective, longitudinal Denver Youth Survey, concluded that:
It seems clear that bullying is related to, and types of bullying are differentially related to, concurrent and future involvement in delinquent assault and other forms of delinquency as well. However… for females, involvement in indirect aggression was often related to the lowest involvement in future delinquency. Also, for both genders, those involved in physical bullying (with or without involvement in indirect aggression) generally had the highest concurrent and future delinquency prevalence rates. (p. 19)

**Internalizing behavior and mental health outcomes**

Emotional and mental health outcomes, including internalizing behavior problems, are also associated with bullying. In particular, depression, anxiety, and anger are frequent correlates for children and youths who bully others (e.g., Baldry, 2004; Kaltiala-Heino et al., 2000; Leadbeater et al., 2006; Roland, 2002; Salmon et al., 1998; Seals & Young, 2003; Slee, 1995). For example, Espelage et al. (2001) assessed bullying over a four-month period in a sample of over 500 middle school students and found that increased levels of impulsivity, anger, and depression were linked with greater levels of bullying over time. A four-year longitudinal study by Kumpulainen and colleagues (1999) found that children involved in bullying had a significantly higher rate of psychiatric symptoms than did other children, at a level high enough to be considered psychologically disturbed. Additionally, Nichols (2004) found that involvement in frequent bullying was associated with higher rates of suicidal ideation (see also Kaltiala-Heino et al., 1999; Kim et al., 2005; Roland, 2002). Bullying has been correlated with loneliness and poor social and emotional adjustment (e.g., Bosworth et al., 1999; Nansel et al., 2001), as well as with sadness (Glew et al., 2005). Baldry (2004), in a study of 661 Italian adolescents aged 11 to 15, found that direct bullying such as hitting and name-calling was unrelated to mental and physical health, but indirect bullying such as spreading rumors and ignoring was a significant predictor of anxiety, depression, and withdrawn behaviors.
**Academic outcomes**

Other correlates of bullying include problems related to schooling, such as lack of school engagement (e.g., Forero et al., 1999; Rigby & Slee, 1993), poor academic achievement (Nansel et al., 2001), and school drop-out (Byrne, 1994). For example, Nansel et al. (2003a) collected data from 930 middle school youths at three time points over a period of two years. After controlling for baseline scores, youth classified as bullies at baseline showed poorer school adjustment than did youths who were not involved in bullying. Research also suggests that bullies do not feel school is a safe place to be (Glew et al., 2005). However, not all studies concur that bullying is related to poor academic achievement. Woods and Wolke (2004) dismiss the theory that frustration and failure in school leads to direct bullying behavior. In a study involving interviews with over 1,000 6- to 9-year-old children, no relationship was found between direct bullying and decreases in academic achievement. Relational bullying, however, predicted higher levels of academic achievement.

**Physical outcomes**

Finally, bullying has been associated with some physical health complaints and symptoms, although not nearly to the level exhibited by victims. For example, Fekkes et al. (2004) studied 2,766 Dutch children aged 9 to 12 years and found that bullies were not at elevated risk for the large majority of the health symptoms studied (compared with non-bullies); the exceptions being higher rates of headaches and bed-wetting. Similarly, a British study involving 8,242 children aged 7-8 years found that children with soiling problems were significantly more likely to be involved in overt bullying than were children without soiling problems (Joinson, Heron, Butler, & von Gontard, 2006). In addition, an Australian study of 3,918 11- to 16-year-old students found that bullies were more likely to experience a larger number of psychosomatic symptoms than were non-bullies, including headache, stomachache, backache, irritability, and difficulty sleeping (Forero et al., 1999).
IV. Bystanders

Bullying incidents usually do not involve a single bully and his or her targeted victim, and the last of the four key participant roles in the bully/victim dynamic is the bystander. Coloroso (2005) describes bystanders as:

…the supporting cast whose role is to aid and abet the bully through acts of commission or omission. They can stand idly by or look away, afraid to step in for fear of becoming a target themselves, or they can actively encourage or join in the bullying. (p. 50)

A Canadian study by Craig and Pepler (1997), in which elementary school children were videotaped on a playground, revealed that children other than the aggressive bully and the victim were involved in 85 percent of bullying incidents. In another observational study, O'Connell, Pepler, and Craig (1999) found that the number of peers present increased the duration of the bullying episodes. Peers physically or verbally joined in the aggression in 21 percent of the cases, passively watched the incident in 54 percent of the cases, and intervened to help the victim in 25 percent of the incidents. The study also found that girls in grades 1-3 and girls in grades 4-6 were more likely to intervene to help the victim than were boys in grades 4-6, while boys in grades 4-6 were more likely to actively join a bully and participate in bullying than were boys in grades 1-3.

There are different ways in which a bystander chooses to intervene (or not to intervene) in a bullying situation, and the choice of intervention approach may affect the outcome of the bullying incident. As such, one current area of research in school bullying prevention involves determining how best to encourage bystanders to intervene when they witness a bully/victim situation. A study by Hawkins, Pepler, and Craig (2001) observed 58 children aged 6 to 12 years during playground bullying incidents, and found peers to be present during 88 percent of the episodes. Peers intervened in 19 percent of the bullying incidents, and were more likely to intervene in same-sex incidents, e.g., girls more frequently intervened when the bully and victim were both female. When peer interventions were directed towards the
victim or both the victim and the bully they were likely to be nonaggressive, but peer interventions directly solely at the bully were more likely to be aggressive.

Frequently, bystanders are afraid to associate with victims for fear of becoming victims themselves, and they may be reluctant to try to stop a bullying incident or report bullying behavior to an adult out of fear of retaliation from the bully. The defending of victims may also be related to a bystander's self-concept; Salmivalli (1998) found that bystanders were more likely to defend victims if the bystander displayed competencies in both physical and social self-concept. Sometimes bystanders are uncertain about how best to intervene without worsening the situation for a victim (Limber, 2002). For example, a study of fourth- to sixth-grade students by Melton and colleagues (1998) found that 38 percent of bystanders reported they "did nothing" when they observed bullying because they felt it was none of their business. Twenty-seven percent of the students felt conflicted about intervening, i.e., they had not helped but felt that they should have, while 35 percent reported that they had tried to help. This lack of action and support of the victim may result in bystander feelings of guilt, anxiety and helplessness, or even apathy and contempt (Coloroso, 2005). Bystanders who encourage the bully or who watch silently end up causing more distress to the victim. Furthermore, this type of behavior helps elevate the bully to a glorified role, and increases the likelihood that other bystanders will become desensitized to future bullying behavior.

Outcomes and Correlates for Bystanders

Those children who are not directly involved in incidents of peer victimization as either bullies or victims may still experience negative effects of bullying. Observing their classmates being harassed, humiliated, or ostracized by peers increases the level of daily anxiety reported by bystanders (Nishina & Juvonen, 2005), in part owing to the fear that the bully may target them next (U.S. Department of Education, 1998). Bullying also detracts from classroom learning for bystanders because it interrupts class sessions, takes up teachers' time, and is a distraction from curriculum objectives. As stated by Chandler and colleagues, "students who must think about avoiding harm at school are diverting energy that should be expended on learning" (Chandler et
Moreover, if teachers and school administrators do not intervene in bullying situations, the school climate begins to foster hostile peer group norms in which students learn that they will not be protected from peer victimization because school personnel are either unable or unwilling to control bullies' behavior (Rigby, 1996).

The harmful impacts of witnessing peer harassment may be long lasting. In a study of 79 students from a university in the Midwest, Janson and Hazler (2004) examined psychological distress in bystanders and victims of repetitive abuse, such as bullying. The subjects completed self-report questionnaires, and participated in interviews about their experiences. During the interviews subjects' physiological reactivity was measured through a heart rate analysis and skin conductance test. The authors found that while the victims recalled greater levels of trauma than did the bystanders, both groups reported similar and unexpectedly high levels of psychological distress, which were confirmed through the physiological measurements. The authors contend that the students' distress levels were akin to the levels of distress exhibited by war veterans who have been clinically diagnosed with posttraumatic stress disorder.

Bystanders to bullying incidents may be granted some small beneficial side effects: Nishina and Juvonen (2005) found that although witnessing a bullying incident increased bystanders' levels of anxiety; it also protected them against increases in humiliation on days that they themselves were bullied. In addition, the authors found that witnessing another student's harassment buffered bystanders against increases in their own negative self-perceptions.

**WHY BULLYING OCCURS – THEORETICAL MODELS**

Many theories have been posited for why bullying occurs, and investigators tend to subscribe to one or more such theories when framing their research. Likewise, bullying prevention strategies and programs use different theoretical foundations in their attempts to improve school bullying problems (Rigby, 2004). As with other social problems that have numerous explanatory theories (e.g., crime, violence), it is likely that no single perspective is
ultimately correct and that a number of theoretical explanations for bullying might be applicable for a particular bully, victim, classroom, or school.

Before beginning a discussion of the various explanatory theories for school bullying, it is useful to review some of the literature in which students are asked why they bully others. Boulton and Underwood (1992), in a study of middle school students in the U.K., found the most common response given by bullies and non-involved students is that victims provoke the bullying themselves. In a similar study, Rigby (2002c) surveyed 1,770 8- to 11-year-old Australian students as to possible reasons for bullying a peer. Rigby found a very similar ordering of reasons offered by boys and girls, with the large majority choosing either "because they annoyed me" or "to get even" (see Table 4). Rigby suggests that respondents were seeking justifications for their behavior, and may also have been selecting the least socially undesirable options from the set of response choices.

### Table 4 Reasons for Bullying

<table>
<thead>
<tr>
<th>Reason</th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Because they annoyed me</td>
<td>68.2%</td>
<td>60.7%</td>
</tr>
<tr>
<td>To get even</td>
<td>64.0%</td>
<td>46.0%</td>
</tr>
<tr>
<td>For fun</td>
<td>16.0%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Because others were doing it</td>
<td>14.0%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Because they were wimps</td>
<td>11.3%</td>
<td>7.0%</td>
</tr>
<tr>
<td>To show how tough one is</td>
<td>11.3%</td>
<td>7.0%</td>
</tr>
<tr>
<td>To get things or money from them</td>
<td>6.1%</td>
<td>4.2%</td>
</tr>
</tbody>
</table>

*Rigby, 2002c, p. 561

In the following section, seven major theories for bullying are discussed.

1) Control and Power

One explanation for bullying is that it results from a desire for control and power, as indicated by the finding that a strong personality characteristic of bullies is dominance or the desire to dominate others (e.g., Bjorkqvist, Ekman, & Lagerspetz, 1982; Olweus, 1993c; Pellegrini & Bartini, 2000a). According to the power/control theory, bullying occurs as a result
of interactions between children with varying degrees of personal power (physical, mental, or social), in which the more powerful child repeatedly feels the desire to oppress the less powerful child (Rigby, 2004). Some research suggests that at least some of the personality differences between bullies and victims have a genetic basis, with children differing along dimensions of being aggressive versus passive, extroverted versus introverted, and low in empathy versus highly sensitive to others’ feelings. The argument for a genetic predisposition is supported by O'Connor, Foch, Todd, and Plomin (1980), who found that identical twins were significantly more alike in their tendency to bully others than were fraternal twins.

2) Social Dominance

Another common explanation for bullying is that it is a consequence of social dominance behaviors among schoolchildren. In other words, bullying is used as a means of establishing a place in the peer social hierarchy (e.g., Bernstein & Watson, 1997; Bjorkland & Pellegrini, 2002; Elinoff et al., 2004; Rigby, 2004), or serves to boost social status within a group of children (e.g., Pellegrini, 1998). According to Rigby (2004), the social dominance perspective is consistent with evolutionary theory, in which domination over others serves to ensure one’s survival in a competitive environment; thus "survival of the fittest."

Some researchers speculate that transitioning from elementary school to middle school and junior high is stressful and may promote bullying behavior as students attempt to define their position in a new, large and diverse social structure (e.g., Betlem, 2001; Espelage et al., 2001; Pellegrini, 2002). As discussed by Pellegrini (2002), adolescents create a dominance hierarchy by competing with one other for available resources (e.g., peer status and heterosexual relationships), with the most dominant individuals obtaining first claim on these resources. For example, a study of 154 fifth-grade students transitioning through the first year of middle school found evidence that bullying enhanced within-group status and popularity (Pellegrini et al., 1999). Likewise, a study by Coie et al. (1991) found that among first-graders, popular boys bullied aggressive boys to elevate their position among peers. However by third-grade, once peer hierarchies were more stable, rejected boys tended to be the bullies while
non-aggressive boys tended to be the victims. Similarly, a study of over 500 students in grades 6 to 8 found an increase in bullying behavior among sixth-graders over a four-month period, but no increase for seventh- or eighth-graders (Espelage et al., 2001), which the authors suggest was owing to sixth-graders attempting to adapt to the first year of middle school. Further support for this premise was found in a study assessing bullying over two years among children transitioning from primary to secondary school (Pellegrini & Long, 2002). The authors observed that bullying increased as children transitioned between schools, but subsequently decreased once the social hierarchy was re-established.

Once the social hierarchy in a group is set, bullying decreases (as long as peer group membership remains stable) because cooperative behavior increases the likelihood that group members will achieve their common goals (Pellegrini, 2002). In a study of 87 boys entering middle school, Pellegrini and Bartini (2001) found positive, significant correlations between aggression and dominance at the beginning of the year, and no significant correlations between dominance and affiliation. At the end of the year the pattern was reversed; affiliation was significantly related to dominance, but aggression was unrelated. In other words, the boys used aggressive strategies to attain dominance at the beginning of middle school, and once dominance was established they began using affiliative strategies in an effort to solidify their positions. For example, the group leaders frequently used aggression in their initial confrontations, but after winning the encounter they attempted reconciliation with the losers to gain their support.

3) Social Learning Theory

Yet another explanatory theory is that bullying is a learned social behavior, and is often the result of the influence of dysfunctional, conflict-laden families or authoritarian, overbearing parents (e.g., Loeber & Dishion, 1984; Rigby, 1994; Schwartz et al., 1997). Albert Bandura’s classic 1960’s "bobo doll" experiments in social psychology found that children, after watching an adult physically abuse a life-sized doll, mimicked the behavior and treated the doll aggressively when left alone with it to play (Bandura, 1977). Bandura's research led to the
development of "social learning theory," which proposes that behavior is learned through observation, imitation, and modeling. In other words, children learn aggressive behaviors, and the consequences of these behaviors, through watching others. If parents or older siblings engage in bullying at home, children are likely to imitate these behaviors – especially if they observe that the actions are rewarded or lead to successful attainment of goals. Examples of a child's learned behavior might include observing their father exerting control over their mother through physical violence or threat of violence, watching an older sibling gain attention with displays of aggressive behavior, or seeing bullying behavior depicted on television. For example, a prospective study by Schwartz et al. (1997) examined the family experiences of 5-year-old boys who at ages nine and ten become both aggressive and bullied in school. The authors found that boys characterized as aggressive or bullied at ages 9 and 10 had experienced greater exposure to adult aggression and conflict in the home at age 5.

Some researchers have suggested that media effects may be partly responsible for the development of bullies. Bullying shown on television and in movies often goes unchecked and the behavior is sometimes rewarded (Coy, 2001). Remboldt (1994), as cited in Ma (2001), suggests that the media portrays aggressive behaviors, including both physical conflicts and coercion, as exciting, sexy, and effective ways to achieve goals. The concept of "an eye for an eye" is also prevalent in film, books, and television as a way of resolving bullying problems (Oliver et al., 1994b). In addition to general violence and aggression, attempts to harm or manipulate others through subtle methods or via scheming (i.e., indirect or relational bullying) are major themes in entertainment (Coyne, 2004). Popular teen movies often incorporate models for such behavior; examples are Heathers (1989), Cruel Intentions (1999), and Mean Girls (2004). It may be that after watching gossiping and rumor spreading as a form of entertainment, teen viewers became more likely to use this form of aggression in real life. Coyne and Archer (2004) conducted a content analysis of 228 hours of television programs popular among British adolescents, including EastEnders, Friends, The Simpsons, and Star Trek. The authors found that indirect aggression was portrayed in 92 percent of the programs, surpassing physical aggression (55 percent) and verbal aggression (86 percent).
perpetrators of the aggressive acts were more likely to be female, attractive, rewarded, justified, and realistically portrayed. In addition, a longitudinal study by Huesmann, Moise-Titus, Podolski, and Eron (2003) reported that the amount of television violence girls viewed at ages 6 to 10 was predictive of their level of indirect aggression in adulthood, while Coyne and Archer’s (2005) study of 11- to 14-year-old British adolescents found that indirectly aggressive girls viewed more indirect aggression on television than did non-aggressive girls.

Another study by Coyne and colleagues used a laboratory design to examine the influence of viewing indirect aggression on subjects' attitudes and behavior (Coyne, Archer, & Eslea, 2004). Study participants were 199 adolescents aged 11 to 14, who were shown a video portraying indirect aggression, physical aggression, or no aggression (non-violent sports clips). Before viewing the videos the participants were purposely angered by an arrogant male confederate posing as a research assistant, who told them they had performed miserably on a standard test, and then stormed out of the room. Following the video, the subjects were asked to evaluate the research assistant, and were informed that their evaluation would have a direct bearing on his chances of being re-hired at the university and being offered a salary raise. Participants were then given a series of vignettes portraying aggressive scenarios, and were asked what they would do if faced with each situation. Results showed that participants who had viewed either of the aggressive videos rated the confederate more negatively (which the authors considered a form of indirect aggression, akin to rumor-spreading) than participants who had viewed the non-aggressive video. Furthermore, when responding to the vignettes, participants who had watched the indirect aggression video listed more indirect ways of harming others, while participants who had watched the physical aggression video listed more physically aggressive ways of responding. Participants who viewed a non-aggressive video listed responses as would be expected by chance. These results, though limited by external validity concerns, suggest that television, movies, and other forms of modern media may increase the display of aggressive and bullying behaviors in adolescents.
4) Negative Peer Influence

A common concern among parents is the effect of negative peer association on their children, especially during late childhood and early adolescence when peer group membership becomes extremely important to youths (Pellegrini, 2002). At this age peer groups tend to form based on similarities in proximity, gender, and race/ethnicity (Hartup, 1996; Rigby, 2004) and research often finds that peer groups are alike in behavior and attitudes for outcomes such as academic achievement (Wentzel & Caldwell, 1997), delinquency (Elliot & Menard, 1996), and smoking (Ennett & Bauman, 1994). With regard to bullying, Espelage et al. (2003), in a study of 6th- to 8th-grade students, found support for the peer similarity hypothesis with respect to bullying over a one-year period. In other words, students tended to be friends with peers who bullied at similar frequencies as they did. In addition, the authors found that youths who spent time with peers who bullied others reported an increase in their own bullying behaviors over the school year, suggesting that peers’ bullying may influence individual bullying levels. Rigby (2004) proposes that “the acts of bullying are seen as typically sustained by a connection with the group (which may be described as peer pressure or allegiance to a group) rather than by individual motives such as personal malevolence” (p. 294-295).

In addition to selecting peers based on similarity, some research indicates that early adolescents are attracted to peers who possess characteristics that reflect independence from adult control, such as aggression, delinquency, and disobedience. For example, Bukowski, Sippola, and Newcomb (2000) studied 217 young adolescents around the time of entry to middle school, and observed that both girls and boys experienced an increase in their attraction to aggressive peers. Similarly, a longitudinal study of middle school students found that when asked to nominate a date to a hypothetical party, girls were marginally more likely to select boys who were considered dominant and aggressive (Pellegrini & Bartini, 2001). This difference emerged despite controlling for physical attractiveness and peer affiliation (i.e., popularity). If students are attracted to aggressive peers who bully others, their own levels of bullying behaviors may increase as a result of this exposure.
5) Socio-Cultural Phenomenon

Rigby (2004) suggests that bullying may be a function of the existence of multiple social groups with differing levels of power. These "levels of power" may refer to such differences as gender, race/ethnicity, socioeconomic status, or religion. Most societies are patriarchal, with males holding more power than females and having a tendency in oppressive behaviors (Rigby, 2004). As discussed previously, the bulk of the research finds that boys are more likely than girls to be bullies. Rigby suggests that cross-gender bullying and aggression may in part reflect the way in which some boys have learned to behave towards girls, i.e., through the expression of dominance. This theory is also useful for explaining how some boys frequently bully and oppress other boys who do not possess stereotypical masculine characteristics such as physical strength, love of sports, and so forth.

Beyond gender, research has not been consistent with respect to the impact of other social power differences on bullying. A number of studies have assessed the relationship between student racial/ethnic background and bullying/victimization, and based on the current evidence bullying is not reliably associated with racial or ethnic differences across a range of geographic locations. For the most part, racial/ethnic concerns are dependent on regional characteristics, particularly with respect to the racial/ethnic distribution of the population residing in a particular region. For this reason, studies on racially-based harassment tend to be specific to the schools and districts in which they are conducted, and are usually not generalizable to other locations. For example, in a study in 418 sixth- and seventh-graders from southern California, Graham and Juvonen (2002) found that African-American students were more likely to be nominated by their peers as aggressive than were Latino and multiethnic students, and were less likely to be nominated as victims. Similarly, Graham, Bellmore, and Mize (2006) found that African-American males were the most likely to be nominated as aggressive by their peers. In contrast, Nansel et al. (2001) found that Hispanic youths reported marginally higher involvement in moderate and frequent bullying than did Caucasian or African-American students. In another study, Hanish and Guerra (2000) assessed 1,956 students in 14 public schools in the Midwest and found that Hispanic children had lower rates
of victimization than did African-American or White children, and that African-American children were less likely than Hispanics or Whites to be the victims of repeat victimization. Other studies have not found race or ethnicity to be significantly associated with peer victimization (e.g., Boulton, 1995; Seals & Young, 2003). Overall, there is no reliable pattern in bullying research with respect to differences in racial/ethnic background that transcends local area characteristics.10

The literature on school bullying also does not reliably conclude that children and youths from lower socioeconomic status (SES) homes are either more or less likely to become bullies or victims than are children from families of higher SES (e.g., Olweus, 1993c). A number of studies report that students from lower SES families experience more bullying and victimization than do students from higher SES families (e.g., Wolke, Woods, Stanford, & Schultz, 2001b). For example, von Rueden et al. (2006) studied a sample of 1,896 8- to 18-year-old students across seven European countries to assess the impact of parental educational status and level of affluence (measured by the family's possession of pre-specified material goods). Results indicate that children and adolescents from lower SES families were more likely to experience bullying than were children and adolescents from higher SES families. Other research fails to confirm the relationship between bullying/victimization and low SES (e.g., Almeida, 1999; Ortega & Mora-Merchan, 1999; Sourander, Helstela, Helenius, & Piha, 2000. Duyme (1990), in a study of 77 adopted French adolescents, found that reports of bullying were not related to social class, however, rates of other antisocial behaviors such as fighting, truancy, disobedience, lying, and destructiveness were higher for youths of lower SES. Different findings were reported by Christie-Mizell (2004), who suggests a U-shaped relationship, with children from both low-income and high-income families at higher risk for bullying than children from middle-class families.

---

10 An exception to this assertion is in the context of bullying based on political events. DeRosier (2004b) assessed positive peer nominations and popularity among fifth-grade students of different ethnic groups at one-month and eight months following the terrorist attacks of 9/11. DeRosier found that only Middle Eastern children showed a large drop in positive peer nominations and popularity over the course of the school year, and the bullying of Middle Eastern and Hispanic students also showed a significant increase over the same period.
6) School Ethos

School ethos is another commonly proposed explanation for bullying behavior. Students are influenced by their perceptions of the school culture, in other words, the broader social context of the school community, which consists of the behaviors and attitudes of students and staff as well as the structural characteristics and functioning of the school itself. For example, a study of middle school boys found that boys who held more positive perceptions of school climate tended to engage in fewer aggressive and delinquent behaviors than did those with more negative perceptions (Kuperminc, Leadbeater, Emmons, & Blatt, 1997). Similarly, a study in a Steiner school that focused on a noncompetitive, group-centered school ethos found a lower level of bullying than typically found at other schools, and no reports of direct physical bullying among students (Rivers & Soutter, 1996). Certain elements of the school ethos have been hypothesized to contribute to bullying, including large classes in secondary schools and students' exposure to a number of different teachers and peers (e.g., Pellegrini et al., 2002), lack of student supervision (e.g., Olweus, 1993c; Pellegrini & Bartini, 2000b; Smith & Sharp, 1994), focus on competition between students and social comparisons between peers (e.g., Akiba, 2001), and school administrator and teacher attitudes toward bullying (e.g., explicit anti-bullying policies versus "It’s a normal part of growing up and kids need to learn how to handle it"; Olweus, 1993c; Smith & Brain, 2000).

School location and size

Contrary to popular opinion, the type of community in which a school is located (i.e., urban, suburban, or rural) is not a consistent predictor of bullying rates. For example, findings by Nansel and colleagues (2001) indicate that victimization rates are similar among youths from urban, suburban, town, and rural areas in the United States. However, small differences were found in bully perpetration rates, with 2 to 3 percent fewer suburban youths participating in moderate amounts of bullying than youths from town, suburban, or urban areas, and rural youths being 3 to 5 percent more likely than others to report that they had ever bullied others. This runs counter to the common belief that bullying is more of a problem for urban schools.
Similarly, a longitudinal study by Kumpulainen et al. (1999) found that the probability of being involved in bullying was higher in rural and semirural areas than in urban areas, particularly for females. The authors suggest that cities might offer advantages such as a larger pool of potential friends, making it easier for children to find others who share their interests and hobbies. In a small town or small school, victims of bullying have fewer friend options.

Class size and school size also appear to be unimportant factors relating to the incidence of bullying (e.g., Olweus, 1993c; Whitney & Smith, 1993). Using several years of Norwegian school data, Olweus (1993c) determined that the incidence of bullying in small, one-room primary schools was very similar and in some cases was even higher than in larger primary schools. Similarly, Boyesen and Bru (1999), in a sample of 1,071 14-year-old Norwegian students, found no difference in the prevalence of peer harassment in small versus large classrooms. However, Ma (2002), in a study of nearly 14,000 students in grades six and eight from 150 from schools in New Brunswick, Canada, found that while school size was not a significant factor in predicting bully victimization, it was a significant factor in predicting bully perpetration. Students in small schools were found to participate in more frequent bullying than were students in large schools.

**Stable group membership**

Pellegrini (2002) proposes that the increased rates of bullying seen in school transition years (e.g., from elementary school to middle school), are in part a result of the breaking up of intact cohorts of friends, because children are less aggressive and more cooperative when they are among familiar peers. Pellegrini suggests that in peer groups with stable membership, aggressive and cooperative acts will result in reciprocal behavior from peers. However, this give-and-take logic only works if the individuals are faced with repeated encounters. Without repeated meetings, there is a lower likelihood that aggressive acts will have the opportunity to be reciprocated, and thus the use of aggression may be instrumental to attain goals or obtain resources. Pellegrini also cites support for the theory of reciprocal altruism via computer simulations based on the Prisoner's Dilemma (Axelrod & Hamilton, 1981), which find that individuals who know that they will meet repeatedly are more likely to behave cooperatively.
than when they know they will meet only once. An experimental study of 192 primary school children also found support for reciprocal altruism (Pellegrini et al., 2002). Participants were asked to engage in four play sessions; pairs of children were either told they would play together for all four sessions, or that they would have a new playmate for each session. Children in the repeated playmate condition were found to resolve significantly more conflicts than those in the new playmate group, suggesting that cooperative behavior was associated with the advance knowledge of repeated interactions.

**Competition between students**

A school environment that encourages and promotes serious competition between students may also be a contributing factor in bullying behavior (e.g., Olweus, 1995). For example, Akiba (2001) examined the effects of high-stakes testing on school violence in Japan, and concluded that students react to this type of environment with increased anxiety about grades and their future, boredom, and devaluation of schoolwork. These outcomes in turn lead them to seek social status among peers at school, which results in collective bullying because of status competition in the peer group.

**Supervision**

Bullying thrives in areas that lack proper adult supervision; as such, higher rates of bullying behavior are found outside the classroom in areas such as the playground, in hallways, in the cafeteria, in locker rooms, in parking lots, and in bathrooms (Craig et al. 2000; Craig & Pepler, 1997; Olweus, 1993c). The lack of supervision in schools may contribute to the bullying problem by increasing the available opportunities for peer harassment (U.S. Department of Education, 1998). Some research finds that up to 75 percent of victims are bullied during recess or lunch periods (Whitney & Smith, 1993). Increasing student supervision and monitoring outside of the classroom environment may reduce the incidence of bullying; for example, Boulton (1994) found that child-reported playground bullying was reduced by 40 to 50 percent after lunch-time supervisors attended bullying awareness training sessions. Similarly, based on data from 700 primary and secondary schools, Olweus (1993a) found a
significant, negative relationship between the occurrence of bullying episodes and the number of adults supervising students during free time.

Teacher awareness and attitudes

Certain teacher and school administrator attitudes towards bullying may also contribute to the bullying problem. Although research focus and media attention in recent years may serve to increase teacher awareness of the potential negative impacts of bullying, previous research has found that teachers are likely to notice and react to direct verbal or physical bullying, but are frequently unaware of or ignore indirect aggression (e.g., Griffin & Gross, 2004). Furthermore, some teachers do not perceive bullying to be a serious problem, and therefore are not resolute in their efforts to discourage it (e.g., O’Moore, 2000; Olweus, 1993c). This attitude may include thinking that bullying occurs infrequently and is limited to physical acts that are usually identified by teachers, believing that students should learn to deal with aggression and harassment on their own, or judging bullying to be a necessary and character-building experience of growing up (e.g., Smith & Brain, 2000). While it is certainly not the case that all teachers hold such attitudes and it is likely that many are well aware of the extent and seriousness of the problem, the laissez-faire attitudes of some teachers may contribute to the low levels of reporting by victimized youths (e.g., Limber, 2002; O’Moore, 2000; Whitney & Smith, 1993).

7) Theory of Mind

One widely cited explanatory model of aggression suggests that impairment in social information processing leads to the development of aggressive behavior (see Dodge & Coie, 1987; Crick & Dodge, 1994). According to Crick and Dodge, aggressive individuals are more likely to mistakenly attribute hostile intent to other's actions, and to display poor understanding of other's mental states. In other words, bullies are powerful, oafish youths who have little understanding of other's feelings or perspectives and carelessly inflict suffering on innocent victims. John Sutton and his colleagues (Sutton et al., 1999a; Sutton et al., 1999b) question whether this social skills deficit model applies to aggressive behavior in the form of bullying,
proposing that because bullying involves many forms of direct and indirect aggression as well as physical violence, it is likely that at least some bullies have a thorough understanding of social behaviors and use this understanding to their advantage. Sutton frames his research using the "theory of mind," which refers to the ability of individuals to explain and predict behavior through attributing mental states and emotions to themselves and to others. A well-developed theory of mind could prove very useful for certain types of bullying behaviors, such as manipulating friendships, organizing gangs, and behaving in covert ways to avoid detection. Sutton and colleagues contend that some bullies possess good theory of mind because they successfully engage in such behaviors; they also purposefully and skillfully target vulnerable children who are likely to make good victims. Sutton, Smith, and Swettenham (1999c) argue:

_We believe it is important to realise that some bullying children do have power, and that they can misuse this power in ways advantageous to them (in some circumstances). For some, this power takes a social rather than physical form, and such bullies are undoubtedly skilled at achieving interpersonal goals. They would probably not see their behaviour as incompetent or maladaptive, and there is evidence that it often is not._ (p. 133)

The theory of mind was tested by Sutton et al. (1999b), a study in which subjects were scored on their levels of involvement in six key roles: bullies, assistants, reinforcers, defenders, outsiders, and victims, and were classified in one primary role. These scores were assessed with regard to the subjects' performance on a set of 11 stories that assessed understanding of cognition and emotions. Sutton et al. found that, after controlling for age and verbal ability, bullies scored significantly higher on the stories than did students in any of the other participant roles. Furthermore, the bully role score was positively correlated with the overall social cognition score, while the victim role was negatively correlated with social cognition, leading Sutton to question the converse proposal: that individuals may become victims because they have low theory of mind skills. Research by Bjorkqvist, Osterman, and Kaukiainen (2000) lends support to this theory through findings that social intelligence has a strong, positive relationship with indirect aggression.
Conclusion on Theoretical Explanations for Bullying

All seven of the explanatory theories have strengths, and no one perspective is consummate in its ability to explain why peer victimization occurs. A conclusive resolution as to the relative importance of the aforementioned theories is worthwhile in so much as it can be used to help design prevention efforts. Each of these theories has strengths and limitations. For example, a view of bullying that focuses on negative peer influence has strengths in that it captures the development of friendships based on propinquity and other similar traits, as it explains how youths tend to behave in ways that are similar to their friends. However, this view does not explain why aggressive behavior might occur within or between friend groups, or how the school environment might encourage or dissuade bullying behavior. Similarly, the theory of mind perspective dispels the stereotypical bully image and characterizes bullies as intelligent, intuitive, and creative individuals, who use their skills strategically to manipulate peers and attain valued goals such as social status. But this framework fails to provide an understanding of why bullies might seek to manipulate others, namely for reasons such as a desire for power or a quest for social dominance, nor does it explain behavior for bullies who do fit the stereotypical oafish role and choose to harass and intimidate victims in less subtle ways. Again, no single explanatory theory encompasses all possible reasons for which an individual chooses to engage in bullying, but all seven theories offer valuable perspectives that might be useful in explaining particular types of bullying situations.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

School bullying is clearly an event so prevalent that it is encountered by nearly every American child who attends school. Cross-cultural research shows that the phenomenon is not limited to the United States or to Western cultures, and the widespread occurrence suggests that to some extent bullying may be the result of basic human nature such as a desire to dominate or to obtain valuable resources. Inherent or not, research over the past 15 years clearly demonstrates that bullying and victimization are linked to many negative outcomes, some of which last beyond the school years and into adulthood.
The increased media and research focus in recent years has uncovered a wealth of information about peer victimization at school, helping to develop a much better understanding of direct and indirect behavior types as well as to identify those children most at risk for involvement as victims, perpetrators, or, most perilously, as bully-victims. It is encouraging that school personnel, researchers, and policymakers have come to recognize that despite its long history bullying is not benign and should not be considered a "normal" part of child development.

It is strongly recommended that future bullying research focus on longitudinal studies using rigorous methodologies for estimating bullying outcomes, to allow for an accurate assessment of the long-term effects of peer victimization and the expected behavior trajectories for bullying and victimized youths. While it is has been well-established that bullying and victimization are related to antisocial, internalizing, and externalizing behaviors, the majority of the existing research is correlational and does not effectively measure the causal influence of bullying on such outcomes.

Developing a thorough understanding of the characteristics of bullies, victims, and bully-victims along with the correlates and risk factors for bullying and victimization is crucial for guiding the design of prevention strategies. Future research on bullying should focus on the development and evaluation of prevention and intervention programs and policies. The identification of effective policies to enhance school safety and an environment conducive to learning is critical; "zero-tolerance" policies adopted by many schools in recent years have been criticized as overly punitive and focused primarily on profiling and labeling (Spivak & Prothrow-Stith, 2001). Schools, parents, and policymakers – not to mention students – all have a vested interest in understanding which programs and policies are proven to be most effective at reducing bullying and victimization in schools. Effective anti-bullying programs will not only reduce the rates of bullying and victimization, but will increase school safety, optimize educational settings for effective learning, and improve the daily social environments of children and youth.
Chapter 2
The Effect of Bully Victimization on Delinquency

INTRODUCTION

School bullying, a subtype of aggression (Griffin & Gross, 2004; Pellegrini, 2002), has emerged as a prominent issue of concern for students, parents, educators, and researchers. The problem is widespread, both in North America and around the world, and the field has developed a sizeable amount of research over the past 15 years (e.g., Craig, Henderson & Murphy, 2000; Rigby, 1998; Salmivalli, Karhunen, & Lagerspetz, 1996b; Sharp & Smith, 1994). While prevalence rates of bullying and victimization vary widely among studies depending in part on research design, definitions of bullying behavior, measurement type, and other such considerations, estimates from large-scale studies are useful in establishing a ballpark for gauging prevalence. In the United States, a nationally representative study of over 15,600 students in grades 6-10 found that nearly 20 percent of the students reported bullying others "sometimes" or more often during the current school term, while approximately 17 percent of the sample reported having been bullied "sometimes" or more frequently (Nansel et al., 2001). In other words, bullying and victimization are common and frequent experiences for millions of students across the country.

The impacts of bullying may be serious for both victims and perpetrators. Hoover et al. (1992) report that over 14 percent of middle- and high-school students are severely traumatized or distressed by bullying, while the National Center for Education Statistics reports that five percent of students either skip school or avoid specific places in school because they are afraid (DeVoe et al., 2005). Prior research has shown that bullying experiences are related to a large range of negative outcomes, including truancy, poor academic achievement, school drop-out, substance use, depression, low self-esteem, eating disorders, anxiety, injury, common health problems, suicide, delinquency, and violence (e.g., Card, 2003; Kaltiala-Heino, Rimpela, Matttunen, Rimpela, & Rantanen 1999; Nishina, Juvonen & Witkow, 2005; Wolke, Woods, Bloomfield, & Karstadt, 2000).
Several hundred empirical studies have been published that allude to the impacts of bullying. The vast majority of these studies employ cross-sectional, quasi-experimental or descriptive designs without rigorous econometric modeling to account for confounding, thereby preventing any reliable claims for causality. Of the studies that do attempt to assess the causal impacts of victimization (or bullying), the majority employ standard regression techniques that may not sufficiently control for differences between the bullied and non-bullied groups.

Outcome studies on victimization predominantly focus on internalizing behaviors such as depression, anxiety, and loneliness, or outcomes such as peer rejection and school achievement (e.g., Hodges & Perry, 1999; Nishina et al., 2005; Olweus, 1994). The current study focuses on the relationship between bully victimization and delinquency, including violence. Of particular interest with respect to school safety is the potential relationship between victimization and extreme violence. For example, a report by the U.S. Secret Service of 37 incidents of school shootings across 26 states found that about 65 percent of school shooters “felt persecuted, bullied, threatened, attacked, or injured by others … a number of attackers had experienced bullying and harassment that was longstanding and severe” (Vossekuil et al., 2000, p. 7). Likewise, an examination by Anderson and colleagues (2001) of all school-associated violent deaths in the United States between July 1, 1994 and June 30, 1999 found that the killers were twice as likely as the victims to have been bullied by peers. These studies present a disturbing picture of the potential impact of bullying on violence, but the descriptive nature of the research designs preclude any causal conclusions from being made. A handful of rigorous studies using inferential models have tested the relationship between victimization and delinquency or aggression, and have found victimization to be positively predictive of delinquency, aggressive behaviors, or both (e.g., Hanish & Guerra, 2002; Nansel, Overpeck, Haynie, Ruan, & Scheidt, 2003b; Rusby et al., 2005; Sourander et al., 2007).
The Current Study

The current study contributes to the sparse literature concerning the impact of bully victimization on delinquency in three ways. First, the use of a large, nationally representative dataset (the National Longitudinal Survey of Youth 1997) allows for conclusions that are most likely generalizable to students across the country. Second, no previous study has assessed the impact of bully victimization on as wide a range of specific delinquency outcomes (10 were examined in the current study) using data that includes multiple waves of a panel survey. Third, the empirical design of this study is more amenable to causal conclusions than is much of the previous research on bully victimization. Specifically, a propensity score matching technique is employed in which bullied and non-bullied subjects are matched on observable baseline characteristics, allowing these characteristics to be eliminated as potential confounders of the estimated effect of bully victimization on delinquency.

The remainder of the paper is organized as follows: Section II presents a description of the data used for the analysis and Section III explains the analytic approach. Section IV describes the key independent variable, the covariates used in the model, and the 10 outcome variables. Section V presents the study findings, including descriptive analyses of the variables, logistic regression estimates, and propensity score matching estimates, and Section VI assesses the robustness of the model's estimate of the impact of bully victimization on delinquency. To conclude, Section VII summarizes the analysis, presents study limitations, proposes future research, and discusses policy implications.

DATA

Overview of the NLSY97

The analysis in this chapter makes use of the National Longitudinal Survey of Youth 1997 (NLSY97) database. The NLSY97 is a longitudinal, nationally representative sample of 8,984 U.S. adolescents who were between the ages of 12 and 16 on December 31, 1996 (i.e., born in 1980-1984). Survey respondents have been interviewed annually since 1997, and at the
time of this writing the questionnaire continues to be administered on an annual basis. Although the primary purpose of the survey is to document youths in their transition from school to the labor force and into adulthood, the NLSY97 is a remarkably rich questionnaire that collects information on a range of topics including employment, schooling and training, income and assets, living environment, health, marital and fertility history, relationships with parents, and behavior such as dating, sexual activity, drug and alcohol use, and criminal activities.

The NLSY97 is sponsored by the Bureau of Labor Statistics, an agency of the U.S. Department of Labor, and is conducted by the National Opinion Research Center at the University of Chicago with assistance from the Center for Human Resource Research at The Ohio State University. Additional funding for the survey has been provided by the U.S. Department of Defense, the U.S. Department of Justice, the U.S. Department of Education, and the National Institute of Child Health and Human Development.

**Sampling**

Two subsamples of survey respondents take part in the NLSY97:

- a cross-sectional sample, designed to be representative of 12- to 16-year-old youths living in the United States during the initial survey round; and
- a supplemental sample, designed to oversample 12- to 16-year-old non-black Hispanic and non-Hispanic black youths during the initial survey round. The purpose of the supplemental sample was to enable more reliable statistical analyses of youths in these racial/ethnic groups.

Respondent households were selected through a three-stage process, involving area probability sampling of primary sampling units (PSUs), segments, and housing units. First, one

---

11 Sampling information was derived from reports by the Center for Human Resource Research (2003a) and Moore, Pedlow, Krishnamurty & Wolter (2000).
hundred PSUs were selected for each of the two subsamples. Second, the PSUs were divided into segments which consisted of single census blocks or clusters of neighboring blocks. A total of 1,151 segments were selected for the cross-sectional sample, while 600 segments were chosen for the supplemental sample. Third, housing units within each segment were selected for screening interviews, resulting in a list of 90,957 residential addresses.

Individual respondents were selected through a two-stage process, which included a screening interview followed by the full questionnaire. Field workers first determined which housing units were occupied and which were vacant, then conducted brief face-to-face screening interviews at the occupied housing units to determine NLSY97-eligibility of individuals in the residence. A total of 75,410 housing units were screened. Eligible youths were aged 12 to 16 years as of December 31, 1996, and listed an in-scope housing unit as their usual place of residence. The pool of eligible youths also included individuals who were temporarily away from their usual residence on vacation, a patient in a general hospital, living at a boarding school, living in a college dormitory, or residing in a jail, prison, or other detention facility. In addition, eligible youths included those with no other home who were staying at an identified housing unit at the time of the screening interview. The sampling universe excluded temporary visitors who normally resided elsewhere, such as in another country or in a psychiatric hospital.

The total target sample size was 10,000 respondents. The cross-sectional screening sample was planned so that it would produce 5,833 interviews with non-Hispanic, non-black youths (a number determined necessary to represent the entire eligible population of American

---

12 For the cross-sectional sample, PSUs in metropolitan areas were defined as Metropolitan Statistical Areas, while PSUs in non-metropolitan areas were defined as single counties or clusters of neighboring counties. For the supplemental sample, PSUs were defined as single counties. Of the 200 PSUs, a total of 53 were selected in both samples. Therefore, the total number of non-overlapping PSUs was 147.

13 For the supplemental sample, segments were classified as having high- or low-density of Hispanic or black youths, as determined by the 1990 Census. Housing units in the high-density segments were sampled at 10 times the rate of housing units in the low-density segments.

14 In the case of youths who could be linked to more than one housing unit, e.g., children of divorced or separated parents, youths were linked to the mother's housing unit (provided the mother was living and her residence was in-scope).

15 While the definition of 'housing units' explicitly excluded group quarters structures (such as prisons, college dormitories, military barracks, and nursing homes), youths living in such quarters were included in the sampling universe through being linked to an in-scope housing unit via their usual place of residence with a parent or guardian.
youths). Likewise, along with respondents obtained in the cross-sectional sample, the supplemental screening sample was planned so that it would yield 1,667 non-black Hispanic youths and 2,500 non-Hispanic black youths. Overall, 9,808 eligible youths were screened as eligible; see Table 5 for details of the sample sizes.16

Table 5 Screening and Target Sample Sizes17

<table>
<thead>
<tr>
<th>NLSY97 eligible youths</th>
<th>N cross-sectional sample</th>
<th>N supplemental sample</th>
<th>N total</th>
<th>Target sample sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-black Hispanic</td>
<td>1,026</td>
<td>1,102</td>
<td>2,128</td>
<td>1,667</td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>1,175</td>
<td>1,371</td>
<td>2,546</td>
<td>2,500</td>
</tr>
<tr>
<td>Non-black/non-Hispanic</td>
<td>5,134</td>
<td>0</td>
<td>5,134</td>
<td>5,833</td>
</tr>
<tr>
<td>Total</td>
<td>7,335</td>
<td>2,473</td>
<td>9,808</td>
<td>10,000</td>
</tr>
</tbody>
</table>

Of the 9,808 eligible youths, 8,984 completed baseline NLSY97 interviews. This included 6,748 youths from the cross-sectional sample, and 2,236 Hispanic and black youths from the supplemental sample (Moore et al., 2000, p. 39). Table 6 lists the sample sizes for the baseline interview by the race/ethnicity and gender of respondents.

Table 6 NLSY97 Round 1 Sample Racial/Ethnic and Gender Composition

<table>
<thead>
<tr>
<th>Round 1 (1997)</th>
<th>Race/ethnicity</th>
<th>Black</th>
<th>Hispanic</th>
<th>Non-black/non-Hispanic</th>
<th>Mixed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td>1,169</td>
<td>977</td>
<td>2,413</td>
<td>40</td>
<td>4,599</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>1,166</td>
<td>924</td>
<td>2,252</td>
<td>43</td>
<td>4,385</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2,335</td>
<td>1,901</td>
<td>4,665</td>
<td>83</td>
<td>8,984</td>
</tr>
</tbody>
</table>

Figure 1 provides a summary of the sampling and selection process for the NLSY97 respondents.

---

16 Because the sampling strategy included all residents of a household unit who were in the appropriate age range, 1,862 households included more than one NLSY97 respondent (usually a sibling). The cohort of respondents thus represents 6,819 unique households.
17 Table 5 adapted from Moore et al. (2000), p. 24.
Figure 1 Selection of NLSY97 Respondents

Phase 1: Selection of Households for Screening

- 200 primary sampling units (147 non-overlapping) were selected from NORC’s\(^a\) 1990 national sample
- 1,751 sample segments were selected from the PSUs
- 90,957 households were chosen for screening from all housing units in the sample segments

Phase 2: Identification of Eligible Respondents

- Screening interviews were completed in 75,410 housing units
- 9,808 youths from the housing units were identified as NLSY97-eligible
- 8,984 of eligible youths participated in the Round 1 NLSY97 survey

\(^a\)NORC = National Opinion Research Center at the University of Chicago

Procedure

During Round 1, both the target youth respondent and one parent received an in-person interview that was approximately 60 minutes in duration. In addition to the interview, a questionnaire administered during the screening process gathered background and demographic information on members of the youth’s immediate family and on his or her household. Subsequent rounds of the NLSY97 survey occurred at approximately one-year intervals. Instead of the full household demographics questionnaire, Rounds 2-7 contained a brief household income update completed by one of the youth's parents. The youth questionnaires were administered by face-to-face interviewers using a laptop and a computer.
assisted personal interviewing (CAPI) system. Questions on sensitive topics such as delinquent or criminal activities were administered with an audio computer-assisted self-interview (ACASI), which enabled youths to answer questions directly into a laptop computer without the interviewer seeing their responses.

After the first survey round, a portion of the initial cohort did not respond to one or more subsequent interviews. Table 7 provides information about response rates by gender and race/ethnicity for each of the seven rounds of the survey used in the current analysis. Overall, response rates were very high – for example, nearly 88 percent of the initial cohort of respondents participated in the Round 5 survey.

21 The CAPI system helps avoid errors and maintains consistency by automatically guiding respondents to question paths and loops depending on their age and their prior responses to survey questions.
<table>
<thead>
<tr>
<th>Gender</th>
<th>Race/ethnicity</th>
<th>Total</th>
<th>Response rate (% retained from Round 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Black</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-black/non-Hispanic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 1: 1997</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,169</td>
<td>2,413</td>
<td>4,599</td>
</tr>
<tr>
<td>Female</td>
<td>1,166</td>
<td>2,252</td>
<td>4,385</td>
</tr>
<tr>
<td>Total</td>
<td>2,335</td>
<td>4,665</td>
<td>8,984</td>
</tr>
<tr>
<td>Round 2: 1998</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,103</td>
<td>2,238</td>
<td>4,283</td>
</tr>
<tr>
<td>Female</td>
<td>1,101</td>
<td>2,095</td>
<td>4,103</td>
</tr>
<tr>
<td>Total</td>
<td>2,204</td>
<td>4,333</td>
<td>8,386</td>
</tr>
<tr>
<td>Round 3: 1999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,062</td>
<td>2,193</td>
<td>4,170</td>
</tr>
<tr>
<td>Female</td>
<td>1,071</td>
<td>2,076</td>
<td>4,039</td>
</tr>
<tr>
<td>Total</td>
<td>2,133</td>
<td>4,269</td>
<td>8,209</td>
</tr>
<tr>
<td>Round 4: 2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,065</td>
<td>2,153</td>
<td>4,117</td>
</tr>
<tr>
<td>Female</td>
<td>1,059</td>
<td>2,027</td>
<td>3,964</td>
</tr>
<tr>
<td>Total</td>
<td>2,124</td>
<td>4,180</td>
<td>8,081</td>
</tr>
<tr>
<td>Round 5: 2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>996</td>
<td>2,110</td>
<td>3,989</td>
</tr>
<tr>
<td>Female</td>
<td>1,036</td>
<td>1,991</td>
<td>3,894</td>
</tr>
<tr>
<td>Total</td>
<td>2,032</td>
<td>4,101</td>
<td>7,883</td>
</tr>
<tr>
<td>Round 6: 2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,033</td>
<td>2,083</td>
<td>3,998</td>
</tr>
<tr>
<td>Female</td>
<td>1,054</td>
<td>1,973</td>
<td>3,889</td>
</tr>
<tr>
<td>Total</td>
<td>2,087</td>
<td>4,056</td>
<td>7,897</td>
</tr>
<tr>
<td>Round 7: 2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,015</td>
<td>2,060</td>
<td>3,929</td>
</tr>
<tr>
<td>Female</td>
<td>1,046</td>
<td>1,916</td>
<td>3,826</td>
</tr>
<tr>
<td>Total</td>
<td>2,061</td>
<td>3,976</td>
<td>7,755</td>
</tr>
</tbody>
</table>

After each round of the survey, NORC survey staff conducted validation re-interviews over the phone with randomly selected respondents. The primary goal of the re-interviews was to confirm that the surveys were administered as reported by the original interviewer.

22 Table 7 adapted from Center for Human Resource Research, 2003a, p. 33.
Sample Weights

In order to achieve unbiased estimates of the population, the NLSY97 data are weighted in each survey year. The weights provide an estimate of how many American adolescents are represented by each survey respondent. Several important sampling biases are likely in a nationally-representative survey like the NLSY97, and the assignment of individual respondent weights allows compensation for:

- **Differences in the selection probabilities of individual cases.** Selection effects may be problematic, since the probability of an individual's participation in the survey is a function of the probability of the individual's housing unit being selected, as well as any sub-sampling (e.g., for race/ethnicity) applied to individuals identified in the screening phase.

- **Subgroup differences in participation rates in the screening phase.** Differences in participation rates for the sample of youths selected for the survey may also affect the generalizability of the sample. For example, respondents in different segments, PSUs, or racial/ethnic groups may have different rates of nonresponse.

Sampling weights for all interviewed respondents are computed by NORC after each survey round, and this weighting corrects for differential non-response in each survey year (e.g., a respondent might skip Rounds 2 and 4, but respond in Rounds 1, 3, 5, 6, and 7).

**ANALYTIC APPROACH**

**Estimating Causal Treatment Effects**

Researchers are often interested in estimating causal treatment effects, in other words, understanding the impact that exposure to some treatment, program or policy has on some outcome of interest for an individual. Another way to think of the effect caused by a treatment

---

23 See Moore et al. (2000) and Center for Human Resource Research (2003a) for a more detailed discussion of the survey weighting process.
is as a missing data problem, since this effect involves comparing the outcome an individual demonstrated under treatment with the potential, but unobserved, outcome the individual would have demonstrated in absence of treatment (Rosenbaum & Rubin, 1983). In other words, a causal effect is measured by the counterfactual – the estimation contrasts what did happen to an individual under one treatment with what would have happened to them under the alternative treatment (Loughran, 2007).

Causal effects are ideally estimated using a true experimental design in which we have confidence that the treated and untreated groups are comparable at baseline. In a true randomized experiment the researcher has control over the assignment of study participants to treatment and control groups. If children were randomly assigned to be bullied or not bullied, then the impact of bullying on delinquency could be evaluated by simply comparing delinquency outcomes for individuals who were bullied with those who were not. Random assignment is critical as it ensures that exposure to treatment is based on chance alone and will not be correlated with the outcome of interest. In addition, randomization in large experiments usually produces comparable, in other words "balanced," treatment groups with respect to observable covariates (Trochim, 2008). Furthermore, since chance alone was used in determining assignment to treatment group, we can be fairly confident that the groups are balanced with respect to unobserved, unmeasured variables as well (including those factors that were not measured as well as those that could not be measured; Loughran, 2007). Accordingly, any pre-treatment differences between the treated and control groups in a randomized experiment are due to chance, therefore, any observed difference in outcomes between the groups represents an unbiased and consistent estimate of the average treatment effect (Loughran, 2007; Shadish, Cook & Campbell, 2002).

However, children are not randomly assigned to bullying exposure, and therefore bullying data is observational as opposed to experimental. Estimation of causal effects in observational data is complicated by concerns of self-selection into treatment, as well as by the potential for pre-existing differences between treated and untreated groups. In other words, estimation of treatment effects using observational data is prone to bias if assignment to
treatment is correlated with characteristics of individuals which may also affect the outcome of interest. If, on average, the individuals who are bullied are also those who tend to engage in delinquent behavior, then direct comparisons of outcomes for bullied versus non-bullied youths are likely to be misleading.

There are two types of selection effects to be concerned with – those effects that are observable, and those that are not observable, but that we have reason to believe may exist (Loughran, 2007). If all of the differences between bullied and non-bullied children were observable (and they were measured), it would be straightforward to control for these overt biases through adjustments such as matching. However, it is difficult to imagine any survey collecting enough information on bullied and non-bullied children to fully measure all of the potentially relevant differences between them. In fact, it is likely that some of the differences are intrinsically unobservable, or very difficult to measure. These are known as hidden biases. For example, previous research suggests that children with poor social skills tend to be bullied at higher frequency than are children with strong social skills (Champion et al., 2003). If poor social skills are also linked with delinquent behavior, then, unless social skills are properly measured by the survey instrument, this construct will emerge as a hidden bias in a multivariate regression (see West, Biesanz, & Pitts, 2000). Since randomization was not used to determine treatment group assignment (bully victimization) for NLSY97 respondents, the treatment and control groups are not only presumed to be nonequivalent at baseline, but are also assumed to have both observable and nonobservable characteristics that are related to treatment group membership.

**Multivariate Regression**

Multivariate regression attempts to deal with potential bias by controlling for observable covariates in the regression equation. One problem with regression estimates is that they cannot control for unobserved variables (or observed variables that are improperly measured). The other problem is that a critical consideration in regression models is functional form – for example, using a linear model implicitly assumes that the linear model is correct.
Regression models are sensitive to model specification such as non-linear transformations and the inclusion of meaningful interaction terms such as bully victimization*sex (see Imbens, 2004). Treatment effect estimates can be susceptible to large biases if they are estimated from simple ordinary least squares regression models or even multilevel regression models (such as random-effects or HLM) that do not sufficiently account for important interactions and nonlinear effects (Macdonald, Stokes, Ridgeway, & Riley, 2007). In the NLSY97 data it is evident that many of the measured attributes of bullied subjects (e.g., sex, English as a second language, child-care experiences, behavioral and emotional problems, substance use, and so forth) differ substantially from those of non-bullied subjects (see Table 11). As a result of these baseline differences, it is unclear whether linear adjustments for these and other observed variables would sufficiently account for potential confounders in order to produce an unbiased estimate of the treatment effect (Macdonald et al., 2007).

**Propensity Score Matching**

Propensity score matching\(^{24}\) is an alternative method to deal with bias that considers the effect of baseline differences in the treatment and control groups. As first described by Rosenbaum and Rubin (1983), the propensity score is *the conditional probability of exposure to treatment given the observed covariates*. In a nutshell, this modeling technique matches individuals having similar sets of pre-treatment observable characteristics and then compares individuals who look alike in all aspects except for their differing treatment status – the "treatment" being a history of bully victimization. In this way, observable covariates are eliminated as potential confounders of the estimated treatment effect, and the data are structured to resemble a randomized experiment (Schonlau, 2006). A major advantage of propensity score matching over traditional regression-based methods is that it is relatively more robust against model misspecification or overfitting. Rosenbaum (1986) contends that the typical analysis of covariance only adjusts for relationships that are linear, while matching adjusts for any

---

\(^{24}\) The method of propensity score adjustment discussed and used here is matching; other propensity score techniques include stratification, using inverse propensity scores as weights (e.g., Schonlau, van Soest, & Kapteyn, 2007), or direct adjustment by including the propensity score in the regression equation.
functional form shown in the relationship between the outcome of interest and the propensity score.

**Variable Selection**

Propensity score matching begins by determining all of the relevant (and observed) covariates and background variables that may be related to selection into the treatment group and to the outcome variable of interest (Berk & Newton, 1985). As with any analysis technique for use with observational data, there may be other non-observed or non-observable predictors of bully victimization and delinquency for which the treatment and control groups differ. In other words, in the absence of random assignment it is impossible to be certain that no confounding variables exist.

Various methods have been proposed to identify the relevant covariates to include in models that effectively adjust for pretest group differences (e.g., see Cochran, 1965; Reichardt, Minton & Schellenger, 1980; West et al., 2000). Using too many variables for propensity score estimation is not a concern, because there is little risk of overfitting the data. Using too few variables is more problematic because it violates the ignorability assumption (the assumption that no important variable is missing). Rosenbaum (2002) outlined the following approach for choosing covariates: Select a preliminary list of variables by using theoretical knowledge of covariates that might be relevant, and conduct exploratory comparisons of these variables in the treatment and control groups. Consider the technique suggested by Cochran (1965), who, based on simulation work, proposed consideration of any covariate for which the \( t \)-value at pretest for the difference between the treatment and control groups was above 1.5. With this preliminary list of covariates, determine the type of adjustment that will be used, and apply an initial run on the variables excluded from the preliminary list. Any of the excluded variables that exhibit a large imbalance between groups after adjustment should be reconsidered. Rosenbaum's method is attractive because it focuses on covariates that are of theoretical relevance, while also considering the level of pre- and post-adjustment balance for these variables.
Matching

After establishing the list of covariates to be used, the next step is determining whether the distributions of the variables are similar for the treated and control groups. If pre-existing differences are found that are related to treatment status, i.e., the covariates are "out of balance," an estimate of the treatment effect may be biased if these differences are also related to the outcome. To account for these overt biases a matching method is used, in which each treated individual is compared with one or more control subjects who have similar baseline covariates. Matching for covariates in this way is called "balancing" the covariates, and is most commonly used when the available sample involves a small treated group and a large pool of potential control subjects (Loughran, 2007).

The matching process does not involve matching each treated subject to a control with the same scores on all covariates. As the number of covariates increases it would become difficult to find matched pairs that are homogenous on all the x variables. Rosenbaum (2002) presents the following example to help explain why this is so: with a list of 20 binary covariates there would be $2^{20}$ possible values of x, meaning over a million possible values for x. Even with a large pool of potential control subjects to select from it might prove impossible to find a control who matches a treated subject on all 20 covariates. The problem is avoided by using propensity scores in the matching process, in lieu of seeking matched sets that are homogeneous in x. A subject's propensity score represents the probability that the subject would have been assigned to the treatment group, given the list of covariates.

If the x variables are balanced, then treatment group status and covariates x are conditionally independent within matched sets. Even though x may strongly predict who will receive treatment, x will no longer predict treatment group assignment for subjects who have propensity scores of the same value. In other words, if two individuals have the same propensity score, the difference in their assignment to treatment is a result of chance (Haviland, Nagin, & Rosenbaum, 2007).

In order to further clarify the concept of the propensity score, the subsequent equations follow descriptions by Rosenbaum and Rubin (1983) and others (Imbens, 2004;
Loughran, 2007; McCaffrey, Ridgeway, & Morral, 2004). These equations make use of counterfactuals to define the treatment effects from observational data of a treated (bullied) subject and an untreated (not bullied) subject.

Consider every individual in the population under study as having two potential values for the outcome of interest. Let

\( y_1 \) designate the individual's outcome with treatment, and
\( y_0 \) designate the individual's outcome without treatment.

Since an individual cannot simultaneously undergo both states of treatment, either \( y_1 \) or \( y_0 \) is observed; it is not possible to observe the individual's counterfactual outcome. The effect of the treatment is measured as the difference in the outcomes \( y_1 - y_0 \). Thus, the expected effect of treatment on an individual who is randomly drawn from the population is defined as

\[
\text{Average Treatment Effect (ATE)} = E(y_1 - y_0).
\]

In this case we are not interested in the ATE so much as the treatment effect on the individuals who actually experienced the bully victimization "treatment". Define a variable \( Z \), such that

\( Z = 1 \) for each individual who has been treated, and
\( Z = 0 \) for each individual who has not been treated.

The average effect of the treatment on the treated is the average effect for those individuals who reported having been bullied. This quantity is defined as

\[
\text{Average Treatment Effect on the Treated (ATT)} = E(y_1 - y_0 \mid Z = 1).
\]
Determining $ATT$ is problematic because we cannot observe counterfactuals. In other words, we cannot observe both $y_1$ and $y_0$ for an individual because half of the data is missing. The naïve method of estimation would be to simply ignore the missing data. Propensity scoring is a more sophisticated approach that imputes an individual's missing data through matching them to the most similar individual who is a member of the group for which the data is missing. The individual's observed outcome, $y$, becomes:

$$y = (1 - Z)y_0 + Zy_1,$$

thus

$$y = y_0 + Z(y_1 - y_0).$$

Because we are using observational data instead of experimental data, unadjusted estimates of the effects of treatment are almost certainly biased because of the differential distribution of $x$ variables across the treatment and control groups. As discussed earlier, "exact" matching of each treated subject to a control subject who appears nearly identical on the $x$ variables is rarely feasible. The alternative is to form sets of treated and control groups who have similar distributions of $x$.

To estimate the propensity score we conduct a logistic or probit regression with treatment group status as the dependent variable, and include in the model all the relevant covariates and background variables (the observed $x$ variables). Specifically,

$$\text{propensity score} = E(x) = P(Z = 1 \mid x).$$

The propensity score may then be used to adjust for covariates prior to estimating the treatment effect. Rosenbaum and Rubin (1983, 1984) demonstrated that matching the

---

25 The logistic model is a type of regression in which the outcome variable is dichotomous: $P(Z = 1 \mid x) = \frac{e^{x^T \beta}}{1 + e^{x^T \beta}}$, and probabilities predicted from the model are restricted to the interval $[0,1]$.

26 Recently, some researchers have used a boosted regression technique instead of a logit or probit model (see McCaffrey, Ridgeway, & Morral, 2004), but this method is not always preferable (Schonlau, 2006).
treatment and control groups on propensity scores minimizes pretest differences between the groups across the set of observed variables.

*Nearest neighbor matching* is a commonly used matching strategy and is employed in the current analysis. One-to-one matching is when each treated subject is paired with exactly one control subject, the match being the individual who looks the most similar with respect to the propensity score. The matching with controls occurs with replacement, therefore one control case can serve as the comparison for numerous treated cases; thus, the order of matching is irrelevant. Instead of one-to-one matching, it is also possible to match each treated subject to two or three or more control subjects, thereby reducing sampling variability. In theory, using multiple controls per treated subject involves more information, which should improve the estimate of the treatment effect by reducing the standard error of the estimate (Loughran, 2007).

**Common Support**

A check for *common support* is frequently used in propensity scoring techniques. Common support means that the distributions of treated and control subjects are largely or fully overlapping. This ensures that one is comparing "apples to apples." To illustrate, suppose all of the individuals who were bullied were from low income households, while all those who were not bullied were from high income households. Then, the effect of household income would confound the effect of bully victimization. Visually, the two groups would represent separate, non-overlapping clouds of data, preventing implementation of matching because the treated and control subjects would not have similar distributions of $x$. Without overlap it is not possible to conduct appropriate comparisons – attempting to estimate the effect of bully victimization on children from upper-income homes would amount to extrapolation. A regular regression would not consider the issue of overlap and would run a line of best fit through the two clouds of data without recognizing that they are separate.

The amount of common support between treated and control observations may be large or small. Suppose we had a continuous range for household income, with the treated
group ranging between $10,000 and $80,000 per year, and the control group’s income ranging from $20,000 to $150,000. In this case the overlap – or common support – would exist for those subjects with household income values between $20,000 and $80,000. A clean comparison would discard subjects with household incomes in the very low ranges and the very high ranges. In propensity score matching this discarding is not done on the x variables; instead, it is done on the propensity score. The technique of propensity score matching thus allows a determination of whether the treated and control subjects are comparable, and allows those subjects who are not comparable to be ignored in the analysis.

Software and Matching Program

This current study uses Stata SE 9.2 and a propensity score matching model with a user-written program in Stata (Abadie, Drukker, Herr, & Imbens, 2004) to estimate the effects of bully victimization on 10 delinquency outcomes. Treatment and control observations are compared using nearest neighbor matching based on the estimated propensity score.

Specifications of the algorithm include the following:

1. *Three matches*: Three matches per treated observation are used. If two or more control observations are equally close to the treated observation being matched, both are used. In other words, the number of matches per treated observation is greater than or equal to three.

2. *Bias-corrected matching estimator*: Abadie and Imbens (2002) show that the simple matching estimator is biased in finite samples when the matching is not exact. The purpose of the bias-corrected matching estimator is to remove some of the bias that remains after implementation of the matching process, by regression-adjusting the results using the original matching variable. See Abadie et al. (2004) for details.

3. *Estimation of heteroskedastic standard errors*: The program estimates heteroskedasticity-consistent standard errors by conducting a second matching process, this time matching observations in the same treatment group to compare variability in the
outcome for observations with similar propensity score values (Abadie et al., 2004).

4. **Inclusion of NLSY97 survey weights:** The sampling weights adjust for differential survey nonresponse and for the oversampling of black and Hispanic respondents; in other words, they provide an estimate of the number of individuals from the eligible population that were represented by each survey respondent. To account for the use of data from multiple waves of the questionnaire, custom weights were created based on the seven rounds of the survey that were used in the analysis. When determining the three closest matches, the program selects the closest observations such that their summed weights are equal to or just exceed three.

**MEASUREMENT**

**Independent (Treatment) Variable**

In the first wave of the survey, respondents were asked a single question that forms the key independent variable in the current analysis: "Before you turned age 12, were you ever the victim of repeated bullying?" (Center for Human Resource Research, 1999).

**Dependent Variables**

For each of the following 10 delinquency outcomes a single dichotomous dependent variable was created. These variables indicate whether or not the respondent had ever

---

27 The custom weights were created for the NLSY97 cohort via a program available on the National Longitudinal Surveys website, www.bls.gov/nls. According to documentation from the Bureau of Labor Statistics, "using these weights provides a simple method for users to correct the raw data for the effects of over-sampling, clustering and differential base year participation" (Zagorsky, n.d., Section 1, para. 1).

28 While this question is not ideal for modeling purposes, it nevertheless represents an assessment of respondent experiences with bully victimization in their pre-teen years. The unfortunate reality is that longitudinal data on bullying is lacking in the field. While some researchers claim that other datasets assess bullying through proxy measures such as fighting, hitting, and yelling at someone, a thorough examination of the definition of bullying suggests that such proxy measures are not sufficiently specific. To be clearly classified as bullying without utilizing the term 'bullying' or 'bullied', the measure must allow for a separation of true bullying behaviors from other types of interpersonal aggression among students. See Chapter 1 of the dissertation for a review of the criteria for labeling true bullying behaviors. In brief, bullying includes malicious intent, repetition, and a power differential between bully and victim.
experienced the outcome or engaged in the behavior in the years following the baseline survey. For example, respondents were asked whether they had been suspended from school. If a respondent answered "yes" to the question in any of waves 2-7 of the survey, he or she is coded as a "yes" to the school suspension variable. If the respondent did not respond "yes" to any wave, but responded "no" at least once (i.e., did not have six waves of missing data), he or she is coded as a "no" for the school suspension variable. The verbatim survey items for the 10 outcome measures follow (Center for Human Resource Research (2000, 2001, 2002, 2003b, 2004, 2005) :

1. **School suspension**
   "Were you suspended from school since [last interview date month/year]?"

2. **Running away from home**
   "Have you run away, that is, left home and stayed away at least overnight without your parent's prior knowledge or permission since the last interview on [date of last interview]?"

3. **Carrying a handgun**
   "Have you carried a hand gun since the last interview on [date of last interview]? When we say hand gun, we mean any firearm other than a rifle or shotgun."

---

29 Wave 1 was not used in the creation of any of the outcome variables due to it serving as the "baseline" year in which the treatment variable, bully victimization, was assessed.

30 The school suspension variable was only asked of respondents who were enrolled in school at the date of last interview and who were living with a parent at the time of the current interview.

31 The running away from home variable was only asked in waves 2-6 of respondents who were under the age of 17 years at the time of the current interview.
4.  **Gang participation**\(^{32}\)

"Have you ever belonged to a gang?" (waves 2-5)

"Have you been a member of a gang since the last interview date on [date of last interview]?" (waves 6-7)

5.  **Selling drugs**

"Since the last interview on [date of last interview], have you sold or helped to sell marijuana (pot, grass), hashish (hash) or other hard drugs such as heroin, cocaine or LSD?"

6.  **Vandalism**

"Since the last interview on [date of last interview], have you purposely damaged or destroyed property that did not belong to you?"

7.  **Theft**

Theft was assessed by a combination of two survey questions – theft of items worth less than 50 dollars, and theft of items worth more than 50 dollars.\(^{33}\) The survey questions were: "Since the last interview on [date of last interview], have you stolen something from a store or something that did not belong to you worth less than 50 dollars?" and, "Since the last interview on [date of last interview], have you stolen something from a store, person or house, or something that did not belong to you worth 50 dollars or more including stealing a car?"

---

\(^{32}\) The NLSY97 questionnaire items for waves 2-5 queried whether the respondent had "ever" belonged to a gang, as opposed to belonging to a gang since "date of last interview". As such, it is possible that gang participation preceded bully victimization. In an attempt to ensure the proper temporal sequence of the variables, gang participation in wave 1 of the survey was tabulated. Those 420 respondents who indicated that they had participated in a gang prior to the baseline survey were dropped from the analysis. Subsequently, when the remaining respondents were asked whether they had "ever" belonged to a gang in waves 2-5, their responses reflected participation subsequent to wave 1 of the survey, and subsequent to any bully victimization identified at baseline.

\(^{33}\) A positive response to either of these items resulted in a "yes" for the theft variable.
8. **Other property crime**

"Since the last interview on [date of last interview], have you committed other property crimes such as fencing, receiving, possessing or selling stolen property, or cheated someone by selling them something that was worthless or worth much less than what you said it was?"

9. **Assault**

"Since the last interview on [date of last interview], have you attacked someone with the idea of seriously hurting them or have had a situation end up in a serious fight or assault of some kind?"

10. **Arrest**

"Since the date of last interview on [date of last interview], have you been arrested by the police or taken into custody for an illegal or delinquent offense (do not include arrests for minor traffic violations)?"

**Control/Matching Variables**

The NLSY97 is a large, rich data set containing a series of questions on individual and family characteristics. Based on the literature with respect to variables that are risk factors for being bullied as well as potential predictors of delinquency, Table 8 defines 20 variables that were identified as important covariates for the regression model and important variables for use in creating propensity scores. This list is arranged into five categories of variables: demographics, early childhood, school, home environment, and behavior:

- **Demographics:** Demographic variables consist of sex, age, census region, race/ethnicity, household size, household income to poverty ratio, and whether the respondent speaks a language other than English at home.

- **Early childhood:** Three variables measure early childhood experiences: the biological mother’s age at the youth’s birth, whether the youth participated in at least 20 hours per week
in child-care from ages 1-5, and whether the youth ever participated in an official Head Start program.

*School* variables include the total number of schools the youth has attended, the size of the youth's current or most recent school, the number of grades ever repeated, and the youth's score on a standardized math test (Peabody Individual Achievement Test).

Measures of *home environment* include indices assessing the level of physical risk and the level of enrichment provided in the home environment, an index quantifying the number of days per week spent in routine family activities, and a scale assessing parents' level of religiosity.

Lastly, the two *behavior* variables consist of an index summarizing youths' self-reports of substance use and a scale assessing behavioral/emotional problems.
**Table 8 Control/Matching Variables (1997 Baseline Survey)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Age</td>
<td>12-18 years</td>
</tr>
<tr>
<td>Census region</td>
<td>Northeast, North central, South, West</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Non-black/non-Hispanic, Black, Hispanic, Mixed race</td>
</tr>
<tr>
<td>Household size</td>
<td>Total number of residents in household</td>
</tr>
<tr>
<td>Household income to poverty ratio&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ratio of household income to poverty level</td>
</tr>
<tr>
<td>Non-English language at home</td>
<td>&quot;Do you now speak any language other than English at home?&quot;</td>
</tr>
<tr>
<td><strong>Early childhood</strong></td>
<td></td>
</tr>
<tr>
<td>Biological mother's age at youth's birth</td>
<td>Biological mother's age when youth was born</td>
</tr>
<tr>
<td>Attended 20+ hours of child-care from ages 1-5</td>
<td>&quot;From ages 1-5, did [name of youth] ever spend 20 or more hours a week in child-care?&quot;</td>
</tr>
<tr>
<td>Participation in Head Start</td>
<td>&quot;Did [name of youth] ever attend an official, government sponsored Headstart program?&quot;</td>
</tr>
<tr>
<td><strong>School</strong></td>
<td></td>
</tr>
<tr>
<td>Number of schools attended&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Number of schools attended (lifetime)</td>
</tr>
<tr>
<td>Current school size</td>
<td>6-category ordinal variable; categories: &lt;100 students, 100-299, 300-499, 500-749, 750-999, 1000+</td>
</tr>
<tr>
<td>Number of grades repeated&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Number of grades repeated (lifetime)</td>
</tr>
<tr>
<td>Standardized math exam score&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Standardized score on the Peabody Individual Achievement Test (PIAT) math exam</td>
</tr>
<tr>
<td><strong>Home environment</strong></td>
<td></td>
</tr>
<tr>
<td>Physical environment risk index&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Index ranging from 0-7</td>
</tr>
<tr>
<td>Enriching environment risk index&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Index ranging from 0-3</td>
</tr>
<tr>
<td>Family routines index&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Index ranging from 0-28</td>
</tr>
<tr>
<td>Parents' religiosity scale&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Scale ranging from 0-6</td>
</tr>
<tr>
<td><strong>Behavior</strong></td>
<td></td>
</tr>
<tr>
<td>Substance use index&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Index ranging from 0-3</td>
</tr>
<tr>
<td>Behavioral/emotional problems scale&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Scale ranging from 0-8</td>
</tr>
</tbody>
</table>

<sup>a</sup>These variables were prepared for the U.S. Department of Labor by Child Trends, Inc. and The Ohio State University (Moore et al., 1999). See Appendix 1 for detailed descriptions of the home environment and behavioral variables.
Treatment of Missing Data

Logistic regression requires complete cases. Preliminary analysis suggested that conducting a complete-case analysis of the dataset would result in nearly two-thirds of the treated subjects being dropped because of missing data on one or more of the x variables. Table 9 displays the missing data across the control/matching variables. For the total 8,833 respondents at baseline, complete responses were available for 9 of the 24 matching variables: sex; age; census regions (Northeast, North Central, and South); race/ethnicity (black, Hispanic, and mixed); and household size. For the remaining 15 variables, missing data ranged from 11 to 4,148 cases.

In order to retain these subjects and allow for a better estimation of the treatment effect, missing data were imputed using the Imputation by Chained Equations (ICE) algorithm implemented in Stata (Royston, 2005b). The ICE algorithm does not limit imputations to integers or values within a pre-determined range. In the current analysis imputed fractions were permitted (see McCaffrey & Elliott, 2007), however out-of-range values were re-coded as the minimum or maximum possible value for the variable based on the existing data. More specifically, out of range values were identified and truncated for the following variables: household income to poverty ratio, biological mother's age at youth's birth, number of schools attended, standardized math PIAT score, family routines index, and the behavioral/emotional problems scale.

34 In other words, the 8,833 subjects who responded to the bully victimization variable at baseline.
35 The dummy variable for census region = West was dropped from the model to enable a full rank matrix.
36 The dummy variable for race/ethnicity = non-black/non-Hispanic was also dropped from the model to allow for a full rank matrix.
37 This imputation program is particularly convenient because it does not require monotone missing value patterns (a requirement of many other procedures). The ICE approach considers the conditional density of a variable given all the other variables in the model, and employs two major steps. The first step imputes a single variable based on multiple regression on a given set of predictor variables. This is performed by ICE's univariate imputation sampling (uvis) program. The regression model used by uvis depends on the type of variable being imputed; i.e., ordinary least squares for continuous variables, logit models for binary variables (such as ordered logit or multinomial logit), and so on. The second step in ICE applies switching regression, an iterative multivariate regression technique, which uses uvis to cycle through the list of variables to be imputed (see van Buuren, Boshuizen, & Mook, 1999). In other words, ICE applies uvis repeatedly in a regression switching mode. For more details on imputation using the ICE program see Royston (2004, 2005a, 2005b).
38 For example, ICE imputed negative values for 36 cases on the "number of schools attended" variable. Since it is not possible to attend fewer than 0 schools, these cases were all re-coded as 0.
Table 9 presents the number of missing cases per variable, and compares the means and standard deviations for the 15 imputed variables before and after imputation. Although slight differences were observed for some of the means after imputation of data, the results were reasonably similar.\(^{39}\) The sensitivity of study outcomes to the imputations was assessed by conducting outcome analyses on a second imputed data set. No noteworthy differences were found with respect to analytic conclusions; results are not presented here.

Table 9 Missing Data for Control/Matching Variables

<table>
<thead>
<tr>
<th>Variable description</th>
<th>N responded</th>
<th>N not responded</th>
<th>Imputed?</th>
<th>Mean before (std. err.)</th>
<th>Mean after (std. err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.49 (.005)</td>
<td>-</td>
</tr>
<tr>
<td>Age</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>14.34 (.016)</td>
<td>-</td>
</tr>
<tr>
<td>Census region - Northeast</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.18 (.004)</td>
<td>-</td>
</tr>
<tr>
<td>Census region - North Central</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.23 (.005)</td>
<td>-</td>
</tr>
<tr>
<td>Census region - South</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.37 (.005)</td>
<td>-</td>
</tr>
<tr>
<td>Black</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.26 (.005)</td>
<td>-</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.21 (.004)</td>
<td>-</td>
</tr>
<tr>
<td>Mixed race</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>.01 (.001)</td>
<td>-</td>
</tr>
<tr>
<td>Household size</td>
<td>8,833</td>
<td>0</td>
<td>No</td>
<td>4.55 (.016)</td>
<td>-</td>
</tr>
<tr>
<td>Household income/poverty ratio</td>
<td>6,455</td>
<td>2,378</td>
<td>Yes</td>
<td>284.27 (3.37)</td>
<td>279.95 (2.87)</td>
</tr>
<tr>
<td>Non-English language at home</td>
<td>7,803</td>
<td>1,030</td>
<td>Yes</td>
<td>.19 (.004)</td>
<td>.20 (.004)</td>
</tr>
<tr>
<td>Biolog. mother's age at birth</td>
<td>8,243</td>
<td>590</td>
<td>Yes</td>
<td>25.52 (.060)</td>
<td>25.49 (.058)</td>
</tr>
<tr>
<td>Attended 20+ hrs child-care</td>
<td>7,698</td>
<td>1,135</td>
<td>Yes</td>
<td>.46 (.006)</td>
<td>.46 (.005)</td>
</tr>
<tr>
<td>Participation in Head Start</td>
<td>7,700</td>
<td>1,133</td>
<td>Yes</td>
<td>.20 (.005)</td>
<td>.20 (.004)</td>
</tr>
<tr>
<td>Number of schools attended</td>
<td>7,792</td>
<td>1,041</td>
<td>Yes</td>
<td>1.45 (.010)</td>
<td>1.46 (.009)</td>
</tr>
<tr>
<td>Current school size</td>
<td>7,760</td>
<td>1,073</td>
<td>Yes</td>
<td>4.88 (.014)</td>
<td>4.87 (.013)</td>
</tr>
<tr>
<td>Number of grades repeated</td>
<td>7,580</td>
<td>1,253</td>
<td>Yes</td>
<td>.19 (.005)</td>
<td>.19 (.005)</td>
</tr>
<tr>
<td>Standardized math PIAT score</td>
<td>5,966</td>
<td>2,867</td>
<td>Yes</td>
<td>96.70 (.253)</td>
<td>95.87 (.208)</td>
</tr>
<tr>
<td>Physical environment risk index</td>
<td>4,685</td>
<td>4,148</td>
<td>Yes</td>
<td>135.42 (2.08)</td>
<td>139.80 (1.53)</td>
</tr>
<tr>
<td>Enriching environment index</td>
<td>5,360</td>
<td>3,473</td>
<td>Yes</td>
<td>1.73 (.011)</td>
<td>1.70 (.008)</td>
</tr>
<tr>
<td>Family routines index</td>
<td>5,329</td>
<td>3,504</td>
<td>Yes</td>
<td>15.05 (.076)</td>
<td>14.39 (.060)</td>
</tr>
<tr>
<td>Parents’ religiosity scale</td>
<td>5,874</td>
<td>2,959</td>
<td>Yes</td>
<td>377.37 (2.08)</td>
<td>375.96 (1.70)</td>
</tr>
<tr>
<td>Substance use index</td>
<td>8,822</td>
<td>11</td>
<td>Yes</td>
<td>1.01 (.012)</td>
<td>1.01 (.012)</td>
</tr>
<tr>
<td>Behav/emotional problems scale</td>
<td>5,361</td>
<td>3,472</td>
<td>Yes</td>
<td>2.14 (.022)</td>
<td>2.30 (.017)</td>
</tr>
</tbody>
</table>

\(^{39}\) While the imputation maintains the covariance structure, it should be noted that the imputed values add no new information to the propensity score estimations over and above the variables with non-missing values.
RESULTS

Descriptive Results for Treatment Variable and Covariates

Table 10 shows the number of treated and control subjects from the total sample of 8,833 respondents – in other words, the number of youths who indicated in the baseline interview (1997) that they had/had not been victims of bullying. Over 19 percent of the sample reported having been victimized before age 12, while nearly 81 percent of youths reported not having being bullied. These findings are fairly consistent with prevalence rates found in other studies on bullying, including the previously discussed nationally representative study by Nansel and colleagues in 2001.

Table 10 Victims of Bullying (1997 Baseline Survey)

<table>
<thead>
<tr>
<th>Victim of bullying before age 12?</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1,713</td>
<td>19.39%</td>
</tr>
<tr>
<td>No</td>
<td>7,120</td>
<td>80.61%</td>
</tr>
<tr>
<td>Total</td>
<td>8,833</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 11 displays selected demographic measures as reported in the 1997 baseline survey for the 1,713 children who reported having been bullied and the 7,120 children who were not bullied. With respect to racial/ethnic background, 21 percent of the black youths reported bully victimization, compared with 20 percent of the non-black/non-Hispanic youths, 18 percent of the mixed race youths, and 16 percent of the Hispanic youths. Males reported substantially higher rates of victimization than did females (22 percent versus 17 percent). Geographically, reports of bully victimization were 20 percent for the North Central, Southern, and Western regions of the country, with a somewhat lower rate of bully victimization in the Northeast (16 percent).
Table 11 Baseline Demographics of Bullied and Non-Bullied Youths

<table>
<thead>
<tr>
<th>Demographic variable</th>
<th>Bullied (n=1,713)</th>
<th>Not bullied (n=7,120)</th>
<th>Total sample (n=8,833)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-black/non-Hispanic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>915 (20%)</td>
<td>3,682 (80%)</td>
<td>4,597</td>
</tr>
<tr>
<td>Black</td>
<td>481 (21%)</td>
<td>1,812 (79%)</td>
<td>2,293</td>
</tr>
<tr>
<td>Hispanic</td>
<td>303 (16%)</td>
<td>1,561 (84%)</td>
<td>1,864</td>
</tr>
<tr>
<td>Mixed race</td>
<td>14 (18%)</td>
<td>65 (82%)</td>
<td>79</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>980 (22%)</td>
<td>3,530 (78%)</td>
<td>4,510</td>
</tr>
<tr>
<td>Female</td>
<td>733 (17%)</td>
<td>3,590 (83%)</td>
<td>4,323</td>
</tr>
<tr>
<td>Census region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>258 (16%)</td>
<td>1,308 (84%)</td>
<td>1,566</td>
</tr>
<tr>
<td>North Central</td>
<td>396 (20%)</td>
<td>1,619 (80%)</td>
<td>2,015</td>
</tr>
<tr>
<td>South</td>
<td>676 (20%)</td>
<td>2,623 (80%)</td>
<td>3,299</td>
</tr>
<tr>
<td>West</td>
<td>383 (20%)</td>
<td>1,570 (80%)</td>
<td>1,953</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>234 (20%)</td>
<td>929 (80%)</td>
<td>1,163</td>
</tr>
<tr>
<td>13</td>
<td>338 (20%)</td>
<td>1,373 (80%)</td>
<td>1,711</td>
</tr>
<tr>
<td>14</td>
<td>344 (19%)</td>
<td>1,493 (79%)</td>
<td>1,837</td>
</tr>
<tr>
<td>15</td>
<td>365 (20%)</td>
<td>1,471 (80%)</td>
<td>1,836</td>
</tr>
<tr>
<td>16</td>
<td>334 (20%)</td>
<td>1,343 (80%)</td>
<td>1,677</td>
</tr>
<tr>
<td>17</td>
<td>96 (16%)</td>
<td>493 (84%)</td>
<td>589</td>
</tr>
<tr>
<td>18&lt;sup&gt;40&lt;/sup&gt;</td>
<td>2 (10%)</td>
<td>18 (90%)</td>
<td>20</td>
</tr>
<tr>
<td>Attended 20+ hours child-care</td>
<td>867 (21%)</td>
<td>3,201 (79%)</td>
<td>4,068</td>
</tr>
<tr>
<td>Speaks non-English language at home</td>
<td>283 (16%)</td>
<td>1,496 (84%)</td>
<td>1,779</td>
</tr>
</tbody>
</table>

<sup>a</sup> The racial/ethnic category "non-Hispanic/non-Black" is used to maintain consistency with NLSY97 documentation. This category consists primarily of Caucasian and Asian respondents.

In the baseline year, NLSY97 survey respondents ranged in age from 12 to 18 years. The rate of bully victimization was around 20 percent for students aged 12 to 16 years. Lower rates of victimization before the age of 12 were reported by 17- and 18-year-olds (16 percent and 10 percent, respectively). Of the 1,779 respondents who indicated that they spoke a language other than English at home, 16 percent reported having been bullied, while 21

---

<sup>40</sup> NLSY97 eligibility rules required respondents to be between the ages of 12 and 16 years as of December 31, 1996. Based on survey responses in the baseline year, 1997, it appears that 20 youths were either 17-years-old as of December 31, 1996, or else they completed the baseline survey sometime after December 31, 1997. Alternatively, coding error may be responsible for the youths aged 18 in the baseline year of the survey.
percent of the youths who had attended some form of child-care for at least 20 hours per week indicated that they had been bullied.

Table 12 presents baseline demographic variables for respondents and non-respondents – in other words, the 8,833 NLSY97 youths who responded to the bully victimization question and the 151 youths who did not. The 151 non-respondents were on average slightly older than the 8,833 respondents (15.17 versus 14.34 years), with statistically significant differences found across most of the age levels. Non-respondents had younger biological mothers (24.29 versus 25.49 years, t=-2.50), and they were considerably less likely to have attended child-care prior to the age of 5 years (35 percent versus 46 percent, \(\chi^2= 6.23\)). Furthermore, non-respondents had repeated more grades in school (0.38 versus 0.18 grades, t= 3.88), and had attended a larger number of schools (1.64 versus 1.45 schools, t= 1.98). Non-respondents were also from lower income families (average household income to poverty ratio of 220.10 versus 279.95, t=-2.87), and they had substantially lower standardized math exam scores (85.21 versus 95.87, t=-5.93).
Table 12 Baseline Demographics of Respondents versus Non-Respondents

<table>
<thead>
<tr>
<th>Demographic variable</th>
<th>Total sample (n=8,833)</th>
<th>Non-respondents&lt;sup&gt;a&lt;/sup&gt; (n=151)</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-black/non-Hispanic&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4,597 (52%)</td>
<td>68 (45%)</td>
<td>$\chi^2 = 2.92$, n.s.</td>
</tr>
<tr>
<td>Black</td>
<td>2,293 (26%)</td>
<td>42 (28%)</td>
<td>$\chi^2 = 0.27$, n.s.</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1,864 (21%)</td>
<td>37 (25%)</td>
<td>$\chi^2 = 1.03$, n.s.</td>
</tr>
<tr>
<td>Mixed race</td>
<td>79 (1%)</td>
<td>4 (3%)</td>
<td>n too small</td>
</tr>
<tr>
<td>Sex</td>
<td>4,510 (51%)</td>
<td>89 (59%)</td>
<td>$\chi^2 = 3.69$, n.s.</td>
</tr>
<tr>
<td>Male</td>
<td>4,323 (49%)</td>
<td>62 (41%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Census region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>1,566 (18%)</td>
<td>19 (13%)</td>
<td>$\chi^2 = 2.71$, n.s.</td>
</tr>
<tr>
<td>North Central</td>
<td>2,015 (23%)</td>
<td>35 (23%)</td>
<td>$\chi^2 = 0.01$, n.s.</td>
</tr>
<tr>
<td>South</td>
<td>3,299 (37%)</td>
<td>60 (40%)</td>
<td>$\chi^2 = 0.36$, n.s.</td>
</tr>
<tr>
<td>West</td>
<td>1,953 (22%)</td>
<td>37 (25%)</td>
<td>$\chi^2 = 0.49$, n.s.</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1,163 (13%)</td>
<td>6 (4%)</td>
<td>n too small</td>
</tr>
<tr>
<td>13</td>
<td>1,711 (19%)</td>
<td>15 (10%)</td>
<td>$\chi^2 = 8.52$, p&lt;0.01*</td>
</tr>
<tr>
<td>14</td>
<td>1,837 (21%)</td>
<td>21 (14%)</td>
<td>$\chi^2 = 4.30$, p&lt;0.04*</td>
</tr>
<tr>
<td>15</td>
<td>1,836 (21%)</td>
<td>41 (27%)</td>
<td>$\chi^2 = 3.64$, n.s.</td>
</tr>
<tr>
<td>16</td>
<td>1,677 (19%)</td>
<td>42 (28%)</td>
<td>$\chi^2 = 7.48$, p&lt;0.01*</td>
</tr>
<tr>
<td>17</td>
<td>589 (7%)</td>
<td>25 (17%)</td>
<td>$\chi^2 = 22.80$, p&lt;0.01*</td>
</tr>
<tr>
<td>18</td>
<td>20 (0.2%)</td>
<td>1 (1%)</td>
<td>n too small</td>
</tr>
<tr>
<td>Attended 20+ hours child-care</td>
<td>4,068 (46%)</td>
<td>35%&lt;sup&gt;c&lt;/sup&gt;</td>
<td>$\chi^2 = 6.23$, p=0.01*</td>
</tr>
<tr>
<td>Speaks non-English language at home</td>
<td>1,779 (20%)</td>
<td>22%&lt;sup&gt;c&lt;/sup&gt;</td>
<td>$\chi^2 = 0.48$, n.s.</td>
</tr>
<tr>
<td>Biological mother's age at youth's birth</td>
<td>25.49</td>
<td>24.29</td>
<td>t=-2.50, p&lt;0.01*</td>
</tr>
<tr>
<td>Household size (mean # individuals)</td>
<td>4.55</td>
<td>4.72</td>
<td>t= 1.24, n.s.</td>
</tr>
<tr>
<td>Household income to poverty ratio</td>
<td>279.95</td>
<td>220.10</td>
<td>t=-2.87, p&lt;0.01*</td>
</tr>
<tr>
<td>Mean # of grades repeated</td>
<td>0.18</td>
<td>0.38&lt;sup&gt;c&lt;/sup&gt;</td>
<td>t= 3.88, p&lt;0.01*</td>
</tr>
<tr>
<td>Mean # of schools attended</td>
<td>1.45</td>
<td>1.64&lt;sup&gt;d&lt;/sup&gt;</td>
<td>t= 1.98, p=0.05*</td>
</tr>
<tr>
<td>Standardized math PIAT score</td>
<td>95.87</td>
<td>85.21</td>
<td>t=-5.93, p&lt;0.01*</td>
</tr>
</tbody>
</table>

<sup>a</sup> Individuals who failed to answer the bullying question in wave 1 out of the total respondent pool of 8,984 youths.
<sup>b</sup> This category consists primarily of Caucasian and Asian respondents.
<sup>c</sup> Percentage of those who responded to the question.
<sup>d</sup> Mean of those who responded to the question.

Unadjusted Results for Dependent Variables

Table 13 presents unadjusted frequency results for the 10 bivariate outcome variables across the total sample of respondents. Again, each of the dependent variables was queried.
during waves 2-6 or 2-7 of the survey, and responses were summed into a 0-1 variable assessing whether respondents had ever engaged in the behavior or experienced the outcome.

Table 13 Unadjusted Results for Dependent Variables across Total Sample

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>N</th>
<th>&quot;Yes&quot; responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>School suspension</td>
<td>8,520</td>
<td>1,992 (23%)</td>
</tr>
<tr>
<td>Running away from home(^{41})</td>
<td>7,513</td>
<td>997 (13%)</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>8,661</td>
<td>1,458 (17%)</td>
</tr>
<tr>
<td>Gang participation</td>
<td>8,214</td>
<td>619 (8%)</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>8,661</td>
<td>1,568 (18%)</td>
</tr>
<tr>
<td>Vandalism</td>
<td>8,662</td>
<td>1,949 (23%)</td>
</tr>
<tr>
<td>Theft</td>
<td>8,663</td>
<td>2,461 (28%)</td>
</tr>
<tr>
<td>Other property crime</td>
<td>8,662</td>
<td>952 (11%)</td>
</tr>
<tr>
<td>Assault</td>
<td>8,662</td>
<td>2,275 (26%)</td>
</tr>
<tr>
<td>Arrest</td>
<td>8,662</td>
<td>1,943 (22%)</td>
</tr>
</tbody>
</table>

Results show that in the total sample, participation in delinquent/criminal behaviors ranged from 8 percent to 28 percent. The most frequently reported behavior was theft, with 28 percent of the sample indicating that they had stolen something at least once over the six waves of measurement. Twenty-six percent of the respondents reporting having attacked or assaulted someone, 23 percent had been suspended from school, 23 percent had engaged in vandalism, and 22 percent reported having been arrested.

Somewhat lower rates of offending were found for selling drugs (18 percent), carrying a handgun (17 percent), running away from home (13 percent), and committing other property crime (11 percent). The lowest rate of self-reported delinquent behavior for the total sample of respondents was for gang participation, with a rate of 8 percent.

Table 14 presents unadjusted results for the 10 outcome variables by treatment group (bully victimization) status. Figure 2 displays the results from Table 14 in a bar chart.

\(^{41}\) The lower response rate for running away from home is in part a reflection of the question only being asked in waves 2-6 of respondents who were under the age of 17 years at the time of the current interview.
Table 14 Unadjusted Results for Dependent Variables by Treatment Group Status

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Bullied youths % of bullied sample</th>
<th>Non-bullied youths % of Non-bullied sample</th>
<th>Between-group difference</th>
<th>Chi-square statistic (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>School suspension</td>
<td>487 (29%)</td>
<td>1,505 (22%)</td>
<td>7%</td>
<td>$\chi^2 = 40.63^*$</td>
</tr>
<tr>
<td>Running away from home</td>
<td>282 (19%)</td>
<td>715 (12%)</td>
<td>7%</td>
<td>$\chi^2 = 55.92^*$</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>378 (22%)</td>
<td>1,080 (15%)</td>
<td>7%</td>
<td>$\chi^2 = 45.91^*$</td>
</tr>
<tr>
<td>Gang participation</td>
<td>161 (10%)</td>
<td>458 (7%)</td>
<td>3%</td>
<td>$\chi^2 = 23.17^*$</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>418 (25%)</td>
<td>1,150 (16%)</td>
<td>9%</td>
<td>$\chi^2 = 62.47^*$</td>
</tr>
<tr>
<td>Vandalism</td>
<td>544 (32%)</td>
<td>1,405 (20%)</td>
<td>12%</td>
<td>$\chi^2 = 113.03^*$</td>
</tr>
<tr>
<td>Theft</td>
<td>643 (38%)</td>
<td>1,818 (26%)</td>
<td>12%</td>
<td>$\chi^2 = 95.92^*$</td>
</tr>
<tr>
<td>Other property crime</td>
<td>272 (16%)</td>
<td>680 (10%)</td>
<td>6%</td>
<td>$\chi^2 = 55.92^*$</td>
</tr>
<tr>
<td>Assault</td>
<td>635 (38%)</td>
<td>1,640 (24%)</td>
<td>14%</td>
<td>$\chi^2 = 138.68^*$</td>
</tr>
<tr>
<td>Arrest</td>
<td>484 (29%)</td>
<td>1,459 (21%)</td>
<td>8%</td>
<td>$\chi^2 = 46.50^*$</td>
</tr>
</tbody>
</table>

*Statistically significant at p< .001

The unadjusted estimates clearly suggest that bullied youths were more likely to commit acts of delinquency than were non-bullied youths – in fact, bullied youths were more likely to report having committed all 10 delinquency outcomes included in the analysis. For example, 38 percent of bullied youths reported having committed an assault, compared with 24 percent of non-bullied youths. Twenty-two percent of bullied youths had carried a handgun compared with 15 percent of non-bullied youths. Similarly, 38 percent of bullied youths compared with 26 percent of non-bullied youths reported stealing something over the course of the survey. Table 14 and Figure 2 show that the between-group differences in rates of delinquency were most pronounced for assault (14 percent), vandalism (12 percent), and theft (12 percent). Moderate group differences were found for selling drugs (9 percent), arrest (8 percent), school suspension (7 percent), running away from home (7 percent), carrying a handgun (7 percent), and committing other property crimes (6 percent). Small between-group differences were found for gang participation (3 percent).
These simple descriptive results do not take into account any pre-existing differences between the groups. For example, the results do not control for whether bullied youths are more likely than non-bullied youths to be male, or to come from more disadvantaged families – both of which are risk factors for delinquency. Nonetheless, these estimates suggest the direction and size of potential treatment effects.

**Adjusted Results Using Logistic Regression**

The equation below depicts results from a basic logistic regression model. Twenty covariates are used in the analysis; these variables are described in detail in Table 8. In brief, the covariates include: sex, age, census region, race/ethnicity, household size, household income to poverty ratio, non-English language spoken at home, biological mother's age at youth's birth, child-care experience from ages 1-5, participation in Head Start, number of schools attended, current school size, number of grades repeated, standardized math PIAT.

---

42 A nonlinear probability model applicable for dichotomous outcome variables.
score, physical risk in the home environment, enrichment in the home environment, time spent in family routines, parent religiosity, substance use, and behavioral/emotional problems.

\[
\text{Delinquency}_{Y1-Y10} = \alpha + \beta_1(\text{Bullied}) + \beta_2(\text{Sex}) + \beta_3(\text{Age}) + \beta_4(\text{Census region Northeast}) + \beta_5(\text{Census region North central}) + \beta_6(\text{Census region South}) + \beta_7(\text{Race/ethnicity black}) + \beta_8(\text{Race/ethnicity Hispanic}) + \beta_9(\text{Race/ethnicity Other/Mixed}) + \beta_{10}(\text{Household size}) + \beta_{11}(\text{Household poverty ratio}) + \beta_{12}(\text{non-English language}) + \beta_{13}(\text{Mother’s age}) + \beta_{14}(\text{Child-care}) + \beta_{15}(\text{Head Start}) + \beta_{16}(\text{# schools attended}) + \beta_{17}(\text{School size}) + \beta_{18}(\text{# grades repeated}) + \beta_{19}(\text{Math score}) + \beta_{20}(\text{Physical environment risk}) + \beta_{21}(\text{Enriching environment}) + \beta_{22}(\text{Family routines}) + \beta_{23}(\text{Parent religiosity}) + \beta_{24}(\text{Substance use}) + \beta_{25}(\text{Behavioral/emotional problems}) + \epsilon
\]

Even though this equation includes a large set of observable characteristics, conclusions regarding the causal effect of bully victimization on delinquency are limited. As with all regression estimations there may still exist unobservables that are correlated with both an individual’s likelihood to become a victim of bullying and his or her delinquency outcomes. Furthermore, this equation represents a classic kitchen sink regression; the covariates were added into a linear model without any concern for functional form. Model misspecification is likely (although not a given), and is certainly cause for concern. This problem is another example of how propensity score matching is useful – an attempt to properly model the effect of bully victimization on delinquency using regression would require much testing for proper model specification, such as the inclusion of interactions or quadratics or splines or other non-linear specifications. Additionally, this type of identification would be required for each of the 10 different outcome variables. In contrast, with propensity score matching the major concern is covariate balancing, which is implemented a single time and is then applicable for use with all outcome variables.

To control for demographic and background characteristics, Table 15 reports odds ratios for the effect of bully victimization on the 10 delinquency outcomes. The first column shows the unadjusted coefficients for the regression of bullying on delinquency outcomes (i.e., without incorporating any covariates in the model). The second column presents the adjusted coefficients for bullying on delinquency outcomes (i.e., incorporating all 20 covariates).
Table 15 Unadjusted and Adjusted Odds Ratios from Logistic Regressions of Bully Victimization on Delinquency Outcomes

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Unadjusted Odds ratio (S.E.)</th>
<th>Z-score</th>
<th>Adjusted Odds ratio (S.E.)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>School suspension</td>
<td>1.48 (.09)</td>
<td>6.35***</td>
<td>1.18 (.08)</td>
<td>2.39*</td>
</tr>
<tr>
<td>Running away from home</td>
<td>1.77 (.14)</td>
<td>7.40***</td>
<td>1.47 (.12)</td>
<td>4.73***</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>1.57 (.11)</td>
<td>6.74***</td>
<td>1.30 (.09)</td>
<td>3.63***</td>
</tr>
<tr>
<td>Gang participation</td>
<td>1.58 (.15)</td>
<td>4.78***</td>
<td>1.41 (.14)</td>
<td>3.39***</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>1.67 (.11)</td>
<td>7.85***</td>
<td>1.30 (.09)</td>
<td>3.71***</td>
</tr>
<tr>
<td>Vandalism</td>
<td>1.88 (.11)</td>
<td>10.53***</td>
<td>1.56 (.10)</td>
<td>6.95***</td>
</tr>
<tr>
<td>Theft</td>
<td>1.74 (.10)</td>
<td>9.72***</td>
<td>1.49 (.09)</td>
<td>6.57***</td>
</tr>
<tr>
<td>Other property crime</td>
<td>1.77 (.14)</td>
<td>7.40***</td>
<td>1.45 (.12)</td>
<td>4.54***</td>
</tr>
<tr>
<td>Assault</td>
<td>1.96 (.11)</td>
<td>11.65***</td>
<td>1.62 (.10)</td>
<td>7.88***</td>
</tr>
<tr>
<td>Arrest</td>
<td>1.52 (.09)</td>
<td>6.79***</td>
<td>1.19 (.08)</td>
<td>2.57 *</td>
</tr>
</tbody>
</table>

*Statistically significant at p<.01, ***Statistically significant at p<.001

Unadjusted odds ratios show that all outcomes are statistically significant at the p<.001 level. In other words, for individuals who were victimized, the odds of being suspended in school are 1.48 times the odds for individuals who were not victimized. The largest effect is found for assault, with bullied youths being 1.96 times more likely than non-bullied youths to attack or assault someone with the intention of causing them harm.

Similar results were found for the adjusted estimates. Although all of the odds ratios decrease in size with the addition of the 20 covariates, the conclusions regarding the estimated effects of bully victimization on delinquency are unchanged. All 10 outcomes show statistically significant treatment effects. These outcomes provide preliminary evidence of a strong, positive relationship between bully victimization and future delinquency after controlling for individual and family characteristics. For example, these results suggest that the odds of a bullied child committing an assault are 1.62 times as high as for a non-bullied child, the odds of committing a theft are 1.49 times as high, and the odds of running away from home are 1.47 times as high.
Propensity Score Estimation

To estimate propensity scores, the first step was a logistic regression of the binary treatment variable (bullied) on the $x$ variables. These results are presented in Table 16.

Table 16 Logistic Regression of Treatment (Bully Victimization) on Covariates

| Victim of bullying before age 12 | Odds ratio (std. error) | Z-score (P<|z|) | 95% C.I. Lower | 95% C.I. Upper |
|----------------------------------|-------------------------|----------------|----------------|----------------|
| Sex (female)                     | .721 (.040)             | -5.85 (0.000)  | .646           | .805           |
| Age                              | .893 (.022)             | -4.60 (0.000)  | .851           | .937           |
| Census region - Northeast        | .744 (.070)             | -3.16 (0.002)  | .620           | .894           |
| Census region - North Central    | .904 (.078)             | -1.17 (0.244)  | .763           | 1.07           |
| Census region - South            | .943 (.075)             | -0.74 (0.462)  | .808           | 1.10           |
| Black                            | 1.02 (.081)             | 0.26 (0.792)   | .874           | 1.19           |
| Hispanic                         | .810 (.087)             | -1.96 (0.051)  | .656           | 1.00           |
| Mixed race                       | .826 (.250)             | -0.63 (0.527)  | .456           | 1.50           |
| Household size                   | .925 (.018)             | -3.96 (0.000)  | .890           | .962           |
| Household income to poverty ratio| 1.00 (.000)             | -1.67 (0.096)  | 1.00           | 1.00           |
| Non-English language at home     | .889 (.094)             | -1.12 (0.264)  | .723           | 1.09           |
| Biolog. mother's age at youth's birth | .993 (.005) | -1.32 (0.188)  | .982           | 1.00           |
| Attended 20+ hours child-care    | 1.17 (.066)             | 2.69 (0.007)   | 1.04           | 1.30           |
| Participation in Head Start      | .981 (.073)             | -0.25 (0.802)  | .848           | 1.14           |
| Number of schools attended       | 1.09 (.043)             | 2.26 (0.024)   | 1.01           | 1.18           |
| Current school size              | 1.03 (.025)             | 1.11 (0.269)   | .979           | 1.08           |
| Number of grades repeated        | 1.09 (.069)             | 1.35 (0.177)   | .962           | 1.23           |
| Standardized math PIAT score     | 1.00 (.002)             | 2.04 (0.041)   | 1.00           | 1.01           |
| Physical environment risk index  | 1.00 (.000)             | 3.14 (0.002)   | 1.00           | 1.00           |
| Enriching environment index      | 1.04 (.042)             | 0.97 (0.333)   | .961           | 1.13           |
| Family routines index            | .997 (.005)             | -0.51 (0.611)  | .987           | 1.01           |
| Parents’ religiosity scale       | 1.00 (.000)             | 1.49 (0.137)   | 1.00           | 1.00           |
| Substance use index              | 1.13 (.031)             | 4.56 (0.000)   | 1.07           | 1.20           |
| Behavioral/emotional problems scale | 1.20 (.022) | 9.76 (0.000)   | 1.17           | 1.24           |

After fitting the regression model, the predicted probability of each individual receiving treatment was estimated. This value represents each respondent’s propensity score.

Figure 3 shows a box-and-whisker plot of the propensity scores for the treated and non-treated groups. The figure shows that there is much overlap in the distribution of scores for the two groups, in other words, there is substantial common support. This suggests that
the groups are suitable for comparison with matching, and we can be confident that we are comparing "apples to apples."

**Figure 3 Box Plot of Propensity Scores for Bullied versus Non-Bullied Youths**

*a* The boxes mark the 25th and 75th quartiles of the propensity scores; the solid lines in the middle of the boxes are the medians. The whisker lines extending above and below the boxes indicate the medians plus and minus 1.5 times the interquartile range (the difference between the 75th and 25th quartile values), respectively. Propensity scores exceeding this range are indicated by the dark circles.

**Covariate Balance Before and After Propensity Score Matching**

The propensity scores are used to adjust for covariates prior to calculating the treatment effect; in the current analysis, this was done by matching. Details of the matching particulars which included 3-to-1 matches, a bias-corrected matching estimator, heteroskedastic standard errors, and inclusion of NLSY97 survey weights were discussed previously (see the section *Sample Weights*). Table 17 presents adjusted and unadjusted means.
for the 24 covariates used in the estimation of the propensity scores. The "Bullied" and "Not Bullied" columns present variable means for the treatment and control groups, respectively, and the "Difference" column shows the difference between the means (Bullied – Not Bullied). The "Effect Size" column presents a Cohen's $d$ for the standardized difference between the means (Cohen, 1988), defined as:

$$d = \frac{\bar{X}_{\text{bullied}} - \bar{X}_{\text{control}}}{s_2}$$

where $s_2$ represents the pooled, unadjusted standard deviation.

Comparing the observed effect sizes before and after propensity score adjustment indicates whether the adjustment is successful in reducing the standardized differences in means between the treated and control groups; in other words, whether the covariates are balanced.
Table 17 Balance of Covariates for Bullied and Non-Bullied Youths Pre- and Post-Adjustment

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Bullied (mean)</th>
<th>Not bullied</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Unadjusted means</td>
<td>Adjusted means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Differen</td>
<td>Effect size</td>
<td>Mean</td>
<td>Differen</td>
<td>Effect size</td>
<td></td>
</tr>
<tr>
<td>Sex (female)</td>
<td>0.43</td>
<td>0.50</td>
<td>-0.08</td>
<td>0.15</td>
<td>0.43</td>
<td>-0.002</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>14.31</td>
<td>14.35</td>
<td>-0.04</td>
<td>0.03</td>
<td>14.37</td>
<td>-0.065</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Census region - Northeast</td>
<td>0.15</td>
<td>0.18</td>
<td>-0.03</td>
<td>0.09</td>
<td>0.15</td>
<td>-0.005</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Census region - North Central</td>
<td>0.23</td>
<td>0.23</td>
<td>0.00</td>
<td>0.01</td>
<td>0.23</td>
<td>0.000</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Census region - South</td>
<td>0.39</td>
<td>0.37</td>
<td>0.03</td>
<td>0.05</td>
<td>0.40</td>
<td>-0.003</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.28</td>
<td>0.25</td>
<td>0.03</td>
<td>0.06</td>
<td>0.27</td>
<td>0.015</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.18</td>
<td>0.22</td>
<td>-0.04</td>
<td>0.11</td>
<td>0.18</td>
<td>0.001</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Mixed race</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.001</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Household size</td>
<td>4.38</td>
<td>4.58</td>
<td>-0.20</td>
<td>0.13</td>
<td>4.42</td>
<td>-0.031</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Household income to poverty ratio</td>
<td>268.28</td>
<td>282.75</td>
<td>-14.48</td>
<td>0.05</td>
<td>264.03</td>
<td>3.420</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Non-English language at home</td>
<td>0.17</td>
<td>0.21</td>
<td>-0.04</td>
<td>0.12</td>
<td>0.16</td>
<td>0.009</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Biolog. mother's age at youth's birth</td>
<td>25.28</td>
<td>25.55</td>
<td>-0.27</td>
<td>0.05</td>
<td>25.29</td>
<td>-0.010</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Attended 20+ hours child-care</td>
<td>0.51</td>
<td>0.45</td>
<td>0.06</td>
<td>0.11</td>
<td>0.50</td>
<td>0.007</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Participation in Head Start</td>
<td>0.21</td>
<td>0.20</td>
<td>0.02</td>
<td>0.04</td>
<td>0.19</td>
<td>0.028</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Number of schools attended</td>
<td>1.50</td>
<td>1.45</td>
<td>0.05</td>
<td>0.06</td>
<td>1.54</td>
<td>-0.038</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Current school size</td>
<td>4.88</td>
<td>4.87</td>
<td>0.01</td>
<td>0.01</td>
<td>4.91</td>
<td>-0.036</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Number of grades repeated</td>
<td>0.22</td>
<td>0.18</td>
<td>0.03</td>
<td>0.08</td>
<td>0.22</td>
<td>-0.001</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Standardized math PIAT score</td>
<td>95.76</td>
<td>95.90</td>
<td>-0.13</td>
<td>0.01</td>
<td>97.05</td>
<td>-1.249</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Physical environment risk index</td>
<td>154.82</td>
<td>136.19</td>
<td>18.63</td>
<td>0.13</td>
<td>154.93</td>
<td>0.450</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Enriching environment index</td>
<td>1.67</td>
<td>1.70</td>
<td>-0.03</td>
<td>0.04</td>
<td>1.67</td>
<td>0.001</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Family routines index</td>
<td>13.92</td>
<td>14.50</td>
<td>-0.58</td>
<td>0.10</td>
<td>13.99</td>
<td>-0.058</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Parents' religiosity scale</td>
<td>376.66</td>
<td>375.79</td>
<td>0.87</td>
<td>0.01</td>
<td>375.09</td>
<td>1.850</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Substance use index</td>
<td>1.23</td>
<td>0.96</td>
<td>0.27</td>
<td>0.24</td>
<td>1.28</td>
<td>-0.056</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Behavioral/emotional problems scale</td>
<td>2.74</td>
<td>2.19</td>
<td>0.54</td>
<td>0.34</td>
<td>2.72</td>
<td>0.019</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

AVERAGE EFFECT SIZE: 0.08 0.02

The Bullied group means are always unadjusted; only the "Not Bullied" control group is affected by the propensity score matching.

Unadjusted, the majority of the covariates were not balanced between groups. For example, pre-adjustment 43 percent of the bullied group was female compared with 50 percent of the non-bullied group. The difference between the groups was -7.63 percent (or -0.08), and the effect size was 0.15. Bullied youths were less likely to come from the Northeast region of

43 Dummy variables for race/ethnicity = non-black/non-Hispanic, and census region = West were dropped from the model to enable a full rank matrix.
the country (15 percent versus 18 percent), they were more likely to be black (28 percent versus 25 percent), and they were less likely to be Hispanic (18 percent versus 22 percent). The mean number of household members was slightly higher for non-bullied youths (4.58 compared with 4.38), while the ratio of household income to the poverty level was somewhat lower (268 versus 283). Bullied youths were less likely than non-bullied youths to speak a language other than English at home (17 percent versus 21 percent) and they were more likely to have spent at least 20 hours per week in child-care from ages 1-5 (51 percent versus 45 percent). Both the average number of grades repeated and the average number of schools attended were also higher for the bullied youths (0.22 versus 0.18 grades repeated; and 1.50 versus 1.44 school attended). The average rating of the youth's physical environment risk in the home was higher for the bullied group (155 versus 136), the rating of enriching home environment was lower (1.67 versus 1.70), and the rating of time spent in family routines was also lower (13.92 versus 14.50). Bullied youths scored higher on the index of substance use (1.23 versus 0.96), as well as the behavioral/emotional problems scale (2.74 versus 2.19).

After implementing matching with propensity scores almost all of the effect sizes decrease – meaning that the treated subjects are matched with controls who appear very similar at pretest. For example, post adjustment the average for sex for both the bullied and non-bullied groups is .43. The difference between groups drops to -0.24 percent, and the effect size for the difference drops below 0.005.

When balancing covariates with propensity scores, the key is to see a decrease in the difference between groups from pre- to post-adjustment. As shown by the average effect sizes for the unadjusted versus the adjusted means in Table 17, the difference between groups does indeed decrease with the implementation of matching. The average effect size pre-adjustment is .08, while the average effect size post-adjustment is .02. This decrease and the small average effect size over all 24 covariates suggest that the variables are well balanced.45

44 As noted previously, 20 covariates were used in the model, however two of these variables (race/ethnicity and census region) were categorical variables implemented with dummies.
Results from Propensity Score Matching

Common Support

Common support was determined for each of the outcome variables; those treated subjects who were off common support were identified and removed from the sample prior to estimation. In other words, treatment observations with propensity scores that are higher than the maximum propensity score of the controls or lower than the minimum propensity score of the controls were dropped from the analysis. Table 18 summarizes the number of subjects off common support for each of the 10 outcome variables.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th># bullied subjects off common support</th>
<th># bullied subjects on support⁴⁶⁺</th>
<th>Total # of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>School suspension</td>
<td>5</td>
<td>1,656</td>
<td>8,520</td>
</tr>
<tr>
<td>Running away from home</td>
<td>5</td>
<td>1,463</td>
<td>7,513</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>4</td>
<td>1,686</td>
<td>8,661</td>
</tr>
<tr>
<td>Gang participation</td>
<td>2</td>
<td>1,538</td>
<td>8,214</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>4</td>
<td>1,685</td>
<td>8,661</td>
</tr>
<tr>
<td>Vandalism</td>
<td>4</td>
<td>1,686</td>
<td>8,662</td>
</tr>
<tr>
<td>Theft</td>
<td>4</td>
<td>1,686</td>
<td>8,663</td>
</tr>
<tr>
<td>Other property crime</td>
<td>4</td>
<td>1,686</td>
<td>8,662</td>
</tr>
<tr>
<td>Assault</td>
<td>4</td>
<td>1,686</td>
<td>8,662</td>
</tr>
<tr>
<td>Arrest</td>
<td>4</td>
<td>1,686</td>
<td>8,662</td>
</tr>
</tbody>
</table>

⁴⁶⁺ As per Table 10, the total number of treated (bullied) subjects was 1,713.

Treatment Effect Estimates

Table 19 presents the estimated treatment effects from the propensity score matching model.⁴⁶ The coefficients indicate the treatment effect for those subjects who experienced bully victimization.

⁴⁵ Numerous attempts were made to decrease all of the post-adjustment effect sizes to below .03 (e.g., for age, black, Head Start, and number of schools attended), such as the inclusion of single and multiple interaction terms in the model, and dropping certain variables that were highly correlated with another variable. None of the modeling techniques was successful in reducing the average effect size to below .02.

⁴⁶ The robustness of these findings with respect to the specifics of the model is examined in Section VI.
Table 19 Propensity Score Matching Estimates

| Dependent variable         | N total | Average treatment effect on bullied (std. error) | Z (p>|z|) | Proportion of bullied who engaged in outcome |
|---------------------------|---------|-------------------------------------------------|----------|--------------------------------------------|
| School suspension         | 8,515   | -0.008 (.012)                                   | -0.68 (p=.497) | .293                                       |
| Running away from home    | 7,513   | .032 (.011)                                      | 2.95**   | .192                                       |
| Carrying a handgun        | 8,657   | -.000 (.011)                                    | -0.02 (p=.982) | .224                                      |
| Gang participation        | 8,212   | .004 (.008)                                     | 0.44 (p=.657) | .105                                      |
| Selling drugs             | 8,657   | .027 (.011)                                     | 2.34*    | .246                                      |
| Vandalism                 | 8,658   | .067 (.012)                                     | 5.56***  | .322                                      |
| Theft                     | 8,659   | .082 (.013)                                     | 6.35***  | .381                                      |
| Other property crime      | 8,658   | .020 (.010)                                     | 2.11*    | .161                                      |
| Assault                   | 8,658   | .067 (.012)                                     | 5.36***  | .375                                      |
| Arrest                    | 8,658   | -.024 (.012)                                    | -1.90 (p=.057) | .286                                     |

*Statistically significant at p< .05; **Statistically significant at p< .01; *** Statistically significant at p< .001

Of the 10 outcome variables assessed, six show significant treatment effects. For example, bullied youths are 3.2 percentage points more likely to run away from home than are youths not victimized by bullies, and they are 2.7 percentage points more likely than non-bullied youths to sell drugs. In other words, after using propensity score matching to adjust for observable group differences, bully victimization is significantly related to an increased likelihood of running away from home and of selling drugs.

Overall results show the largest treatment effects for theft, followed by assault and vandalism, with smaller effects found for running away, selling drugs, and other property crimes. The magnitude of the effect may be thought of in the following way: Of the 19 percent of bullied youths who ran away from home, 3.2 percentage points or approximately 50 cases of running away are a result of being bullied. Although this figure is not exact, it represents a ballpark estimate of the number of youths reporting delinquent behavior that is attributable to their history of bully victimization. See Table 20.
Table 20 Magnitude of Effects for Significant Outcome Variables

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>N total</th>
<th>N treated (bullied)</th>
<th>% of treated who engaged in outcome</th>
<th>Increase in % due to being bullied</th>
<th>% (N) of treated who would have engaged in outcome if not bullied</th>
<th># cases due to being bullied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running away</td>
<td>7,513</td>
<td>1,468</td>
<td>19% (n=282)</td>
<td>3.2%</td>
<td>15.8% (n=232)</td>
<td>n= 50</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>8,657</td>
<td>1,689</td>
<td>25% (n=418)</td>
<td>2.7%</td>
<td>22.3% (n=377)</td>
<td>n= 41</td>
</tr>
<tr>
<td>Vandalism</td>
<td>8,658</td>
<td>1,690</td>
<td>32% (n=544)</td>
<td>6.7%</td>
<td>25.3% (n=428)</td>
<td>n=116</td>
</tr>
<tr>
<td>Theft</td>
<td>8,659</td>
<td>1,690</td>
<td>38% (n=643)</td>
<td>8.2%</td>
<td>29.8% (n=504)</td>
<td>n=139</td>
</tr>
<tr>
<td>Other property crime</td>
<td>8,658</td>
<td>1,690</td>
<td>16% (n=272)</td>
<td>2.0%</td>
<td>14.0% (n=237)</td>
<td>n= 35</td>
</tr>
<tr>
<td>Assault</td>
<td>8,658</td>
<td>1,690</td>
<td>38% (n=635)</td>
<td>6.7%</td>
<td>31.3% (n=529)</td>
<td>n=106</td>
</tr>
</tbody>
</table>

Similarly,

- Of the 25 percent of bullied youths who reporting having sold drugs, 2.7 percentage points or approximately 41 cases are attributable to bully victimization.
- Of the 32 percent of bullied youths who vandalized, 6.7 percentage points or 116 cases are a result of being bullied.
- Of the 38 percent of bullied youths who committed theft, 8.2 percentage points (139 cases) are due to bully victimization.
- Of the 16 percent of bullied youths who committed other property crimes, 2 percentage points (35 cases) are due to having been victimized by bullies.
- Of the 38 percent of bullied youths who reported committing assault, 6.7 percentage points or approximately 106 cases are attributable to bullying.

No significant treatment effects were found for the remaining four delinquency outcomes. This suggests that after comparing bullied youths to non-bullied youths with similar propensity scores – in other words, to control subjects who are comparable at baseline on a large number of covariates – victimization is not related to an increased likelihood of school
suspensions, carrying a handgun, gang participation, or arrests. In fact, although none of the estimates were statistically significant, negative coefficients were found for the school suspension, carrying a handgun, and arrest variables.

Comparison of Regression versus Propensity Score Matching Estimates

Table 21 presents a comparison of treatment effect estimates for the 10 outcome variables using the traditional regression model and the propensity score model.\footnote{Treatment effect estimates for the two models in Table 21 are expressed on a probability scale. Earlier, Table 16 presented the traditional regression model results as odds ratios, and these have been converted to coefficients to allow for a more straightforward comparison with the propensity score estimates.} Results show that the estimates for the traditional model are larger than the propensity score estimates for 9 of the 10 outcome variables; the exception being theft, for which the treatment effect estimates of .082 in the two models are nearly equivalent. Four variables are found to be statistically significant with the traditional regression model but not with the propensity score model: school suspension, carrying a handgun, gang participation, and arrest.

The differences in statistical significance between the traditional and propensity score models may be attributable to one of two things: either the magnitude of the effect size actually changes under the propensity score model, or the propensity score model does not have as much power to detect an effect of the same magnitude. It is clear from Table 21 that the difference in this analysis is due to a difference in the magnitude of the effect size estimated from the two models. More specifically, the estimate for school suspensions in the traditional model is .027 compared with -.008 for the propensity score model. While .027 would not be considered a large effect size according to Cohen's standards (Cohen, 1988), it is substantially larger (.035) than the effectively nil estimate from the propensity score model. Similarly, the effect for carrying a handgun is .038 higher in the traditional regression model than the estimate from the propensity score model (.038 versus -.000). For gang participation the difference in estimates between the two models is .024, while the largest disparity in estimates is for arrest, with a between-group difference of .053.
Table 21 Comparison of Regression and Propensity Score Matching Estimates

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>N</th>
<th>Average treatment effect on bullied</th>
<th>Propensity treated engaged in</th>
<th>Propensity score Model</th>
<th>Average treatment effect on bullied</th>
<th>Propensity treated engaged in outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Z (p&gt;</td>
<td>z</td>
<td>)</td>
<td>Z (p&gt;</td>
<td>z</td>
</tr>
<tr>
<td>School suspension</td>
<td>8,515</td>
<td>.027 2.39*</td>
<td>.292</td>
<td>-0.08</td>
<td>-0.68</td>
<td>.293</td>
</tr>
<tr>
<td>Running away</td>
<td>7,513</td>
<td>.049 4.73***</td>
<td>.187</td>
<td>.032</td>
<td>2.95**</td>
<td>.192</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>8,657</td>
<td>.038 3.63***</td>
<td>.224</td>
<td>-0.00</td>
<td>-0.02</td>
<td>.224</td>
</tr>
<tr>
<td>Gang participation</td>
<td>8,212</td>
<td>.028 3.39***</td>
<td>.111</td>
<td>.004</td>
<td>0.44</td>
<td>.105</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>8,657</td>
<td>.040 3.71***</td>
<td>.248</td>
<td>.027</td>
<td>2.34*</td>
<td>.246</td>
</tr>
<tr>
<td>Vandalism</td>
<td>8,658</td>
<td>.081 6.95***</td>
<td>.322</td>
<td>.067</td>
<td>5.56***</td>
<td>.322</td>
</tr>
<tr>
<td>Theft</td>
<td>8,659</td>
<td>.082 6.57***</td>
<td>.381</td>
<td>.082</td>
<td>6.35***</td>
<td>.381</td>
</tr>
<tr>
<td>Other property crime</td>
<td>8,658</td>
<td>.041 4.54***</td>
<td>.161</td>
<td>.020</td>
<td>2.11*</td>
<td>.161</td>
</tr>
<tr>
<td>Assault</td>
<td>8,658</td>
<td>.096 7.88***</td>
<td>.376</td>
<td>.067</td>
<td>5.36***</td>
<td>.375</td>
</tr>
<tr>
<td>Arrest</td>
<td>8,658</td>
<td>.029 2.57**</td>
<td>.287</td>
<td>-0.02</td>
<td>-1.90</td>
<td>.286</td>
</tr>
</tbody>
</table>

* Statistically significant at p< .05; **Statistically significant at p< .01, ***Statistically significant at p< .001

ASSESSING THE ROBUSTNESS OF FINDINGS

Changes to Propensity Score Matching Model

In any modeling exercise the sensitivity of findings with respect to minor changes in the model parameters or algorithm should be assessed. A number of alternative choices are considered in this context to verify the results presented in the previous section. These include:

- One-to-one matching instead of three matches per treated subject
- Not using sampling weights
- Not using the bias-adjusted estimate
- Assuming homoskedastic standard errors
Table 22 presents differences for these five model options for the outcome variable of assault. In the interest of space, the results for only one of the 10 outcome variables is presented; assault was chosen because it has one of the largest treatment effect estimates found in the primary model.

**Table 22 Propensity Score Model Options for Outcome of Assault**

<table>
<thead>
<tr>
<th>Model #</th>
<th>Model options</th>
<th>Average treatment effect on bullied (std. error)</th>
<th>Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No options (i.e., 1-to-1 matching, no sampling weights, no bias adjustment, homosked. std errors)</td>
<td>.078 (.018)</td>
<td>Z=4.44*</td>
</tr>
<tr>
<td>2</td>
<td>1-to-1 matching; sampling weights</td>
<td>.067 (.019)</td>
<td>Z=3.56*</td>
</tr>
<tr>
<td>3</td>
<td>3-to-1 matching; sampling weights</td>
<td>.067 (.019)</td>
<td>Z=3.56*</td>
</tr>
<tr>
<td>4</td>
<td>3-to-1 matching; sampling weights, heterosked. std errors</td>
<td>.067 (.012)</td>
<td>Z=5.36*</td>
</tr>
<tr>
<td>5^</td>
<td>3-to-1 matching; sampling weights, bias adjustment, heterosked. std errors</td>
<td>.067 (.012)</td>
<td>Z=5.36*</td>
</tr>
<tr>
<td>6</td>
<td>3-to-1 matching, bias adjustment, heterosked. std errors (no sampling weights)</td>
<td>.090 (.014)</td>
<td>Z=6.23*</td>
</tr>
</tbody>
</table>

* Model 5 is the model used in the primary analysis
* Statistically significant at p< .001

Compared with Model 5, the model used in the primary analysis, the treatment effect estimate increases slightly in models 1 and 6 and decreases slightly in models 2 and 3. Not incorporating sampling weights results in a larger estimated treatment effect, and not assuming heteroskedastic standard errors results in a smaller estimated treatment effect. Regardless, the differences among estimates are not large and suggest that model specification does not have a large impact on the analysis conclusions. In all six models, bully victimization significantly predicts future assault.

Further modeling options were tested using an alternative propensity score matching program in Stata (psmatch2; see Leuven & Sianesi (2003) for more details). This program enables options such as the specification of a caliper for the maximum distance of the control subjects from the treated subjects, the use of multiple neighbors for matching (i.e., 3-to-1
versus 1-to-1), radius matching in which matching is performed within a specified radius given by the caliper, and Mahalanobis matching. Table 23 presents these four model options used for the outcome variable of assault.

Table 23 Additional Model Options for Outcome of Assault

<table>
<thead>
<tr>
<th>Model #</th>
<th>Model options</th>
<th>Treated</th>
<th>Controls</th>
<th>Difference</th>
<th>SE</th>
<th>t-statistic</th>
<th>N; # off common support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmatched</td>
<td>.376 .235 .141 .012 11.87*</td>
<td>.376 .297 .079 .018 4.47*</td>
<td>8,658; 4 off support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Nearest neighbor 1-to-1 matching; caliper .01</td>
<td>.376 .284 .091 .015 6.20*</td>
<td>8,658; 4 off support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Nearest neighbor 3-to-1 matching; caliper .01</td>
<td>.375 .280 .095 .013 7.21*</td>
<td>8,658; 4 off support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Radius matching; caliper .01</td>
<td>.375 .297 .078 .018 4.43*</td>
<td>8,661; 1 off support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Mahalanobis matching; caliper .01</td>
<td>.375 .297 .078 .018 4.43*</td>
<td>8,661; 1 off support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Statistically significant at p< .05

Unmatched, the difference between the treatment and control groups is 14.1 percent, with a significant t-statistic of 11.87. Four model options were tested, with results showing group differences ranging from 7.9 percent to 9.5 percent. All four models have significant t-statistics ranging from 4.43 to 7.21, which provides further evidence that the finding for the effect of bully victimization on assault is robust.

Changes to Dependent Variable Measurement

One concern with the current analysis is the proximity of the bully victimization experience to the development of delinquent behavior. If bullying is supposedly a key driver for these outcomes, then the closer it occurs (chronologically) the more plausible the claim for a causal effect. Although a true longitudinal analysis is not conducted herein, the issue of timing of the outcome variables is assessed by using truncated dependent variables, specifically, variables limited to waves 2-4 of the survey instead of waves 2-7. This sensitivity analysis serves as a robustness check to determine whether the results hold using outcome data that are
closer in time to when the respondents reported being victimized. If results of this assessment show that eliminating waves 5-7 result in substantially smaller treatment effects, the inferences from the assessment of bully victimization on delinquency will be less convincing.

Table 24 recreates the analysis from Table 19, substituting the truncated outcome variables into the propensity score matching model. As outlined in Table 18, those treated subjects who were off common support were identified and removed from the sample prior to estimation. The coefficients indicate the treatment effect on those subjects who experienced bully victimization.

| Dependent variable       | N total | Average treatment effect on bullied (std. error) | Z (p>|z|) | Proportion treated engaged in outcome |
|--------------------------|---------|-------------------------------------------------|--------|------------------------------------|
| School suspension        | 8,425   | -.005 (.012)                                    | -0.43 (p=.664) | .275 |
| Running away             | 7,470   | .023 (.011)                                     | 2.17* (p=.030) | .178 |
| Carrying a handgun       | 8,563   | -.010 (.010)                                    | -1.02 (p=.309) | .156 |
| Gang participation       | 8,127   | .003 (.007)                                     | 0.45 (p=.652) | .083 |
| Selling drugs            | 8,565   | .012 (.010)                                     | 1.16 (p=.245) | .184 |
| Vandalism                | 8,565   | .057 (.012)                                     | 4.89*** (p<.001) | .278 |
| Theft                    | 8,567   | .065 (.012)                                     | 5.22*** (p<.001) | .321 |
| Other property crime     | 8,563   | .021 (.010)                                     | 2.31* (p=.021) | .128 |
| Assault                  | 8,565   | .065 (.012)                                     | 5.31*** (p<.001) | .309 |
| Arrest                   | 8,128   | .003 (.007)                                     | 0.45 (p=.651) | .189 |

*Statistically significant at p<.05; ***Statistically significant at p<.001

Of the 10 outcome variables, five show significant treatment effects. Bullied youths are 2.3 percentage points more likely to run away from home than are youths not victimized by bullies, they are 5.7 percentage points more likely to commit vandalism, and they are 6.5 percentage points more likely to report having committed theft. Youths who experienced bully victimization are also 2.1 percentage points more likely to commit other property crimes, and they are 6.5 percentage points more likely to attack or assault someone.
Table 25 compares the treatment effect estimates using the 6-wave versus 3-wave outcome variables. Compared with the original results using six waves of outcome data, four of the truncated dependent variables result in slightly smaller treatment effect estimates (between-estimate difference for running away = -.009, vandalism = -.010, theft = -.017, and assault = -.002), while the estimate for other property crime increases by .001. The estimate for selling drugs drops from .027 to .012, and is no longer statistically significant. The fact that the significant treatment effect estimates hold for five out of six of the outcome variables suggests that the propensity score matching analysis presented earlier is robust.

Table 25 Treatment Effect Estimates for 6-Wave versus 3-Wave Dependent Variables

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Original dependent variables:</th>
<th>Truncated dependent variables:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average treatment effect on bullied (std. error)</td>
<td>Average treatment effect on bullied (std. error)</td>
</tr>
<tr>
<td>Running away</td>
<td>.032 (.011)**</td>
<td>.023 (.011)*</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>.027 (.011)*</td>
<td>.012 (.010)</td>
</tr>
<tr>
<td>Vandalism</td>
<td>.067 (.012)***</td>
<td>.057 (.012)***</td>
</tr>
<tr>
<td>Theft</td>
<td>.082 (.013)***</td>
<td>.065 (.012)***</td>
</tr>
<tr>
<td>Other property crime</td>
<td>.020 (.010)*</td>
<td>.021 (.010)*</td>
</tr>
<tr>
<td>Assault</td>
<td>.067 (.012)***</td>
<td>.065 (.012)***</td>
</tr>
</tbody>
</table>

*Statistically significant at p< .05; **Statistically significant at p< .01; ***Statistically significant at p< .001

DISCUSSION

Bullying is so common a phenomenon that it is almost ubiquitous in American schools. Indeed, the findings by Nansel et al. (2001) suggest that well over 16 million students in the U.S. are currently involved in bullying to some capacity. Until fairly recently bullying was viewed as a normal developmental experience, or rite of passage, for children in schools. Smith and Brain (2000) suggest:

"We can therefore think of bully-victim relationships as normative, in the strictly limited sense that they are likely to be found in any relatively enduring human group that it is difficult for someone to leave if they are

48 Based on 2007 school enrollment data provided by the United States Census Bureau (2007).
...experiencing victimization…. Research to date suggests we can accept as a reasonable generalization that any school can anticipate bullying occurring, although with varying degrees of severity” (p. 2).

Since the mid-1990s a surge of research on bullying has emerged both in the U.S. and in many countries around the world (such as the United Kingdom, Japan, Canada, Scandinavia, Australia, and Italy). The resulting evidence on the potentially serious outcomes for both victims and bullies has alerted researchers, educators, and the public to the possibility of non-trivial negative repercussions of bullying and victimization. Although the prevalence and nature of school bullying and its short term correlates have been studied extensively, little research has examined the potential influence of bullying on delinquency. In particular, very few prior studies have considered the impact of victimization on the development of delinquent behaviors.

This study uses panel data from the NLSY97 to explore the relationship between victimization prior to the age of 12 years and 10 delinquency outcomes measured over a period of six years after the baseline survey. To identify the causal effects of bully victimization on delinquency a propensity score matching technique is employed – a robust approach for producing suitable comparison groups when using observational data. In comparison with multivariate logistic regression results, propensity score matching estimates provide more conservative estimates of the treatment effect.

Using traditional logistic regression and controlling for 20 covariates, bully victimization is significantly predictive of 10 out of 10 delinquency outcomes. After implementation of propensity score matching, results still clearly suggest that early bully victimization is predictive of future delinquency. Of the 10 delinquency outcomes assessed, 6 are found to be significantly related to prior bully victimization. Students who were victimized by bullies are significantly more likely to run away from home, sell drugs, vandalize, steal, commit other types of property crimes, and engage in violent behavior in the form of assault. The treatment effect estimates are not significant for school suspension, carrying a handgun, gang participation, or arrest. Thus, the estimates for these four variables are likely improperly
estimated using the traditional regression model. As summarized in Table 26, the use of propensity score matching finds that once the treated subjects were matched with appropriate comparison subjects, these outcomes no longer appear to be significantly affected by prior bully victimization experiences.

### Table 26 Conclusions for Logistic Regression and Propensity Score Matching Estimates

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Logistic regression – significant?</th>
<th>Propensity score matching – significant?</th>
<th>Difference in conclusion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>School suspension</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Running away</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Carrying a handgun</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Gang participation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Vandalism</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Theft</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Other property crime</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Assault</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Arrest</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

This study provides evidence that early victimization contributes to the development of delinquency and violence. The concept of violence begetting violence is not new (e.g., Curtis, 1963), and studies by Widom have found a positive link between childhood victimization and future violent offending (e.g., Rivera & Widom, 1990; Widom, 1989). What might explain the relationship between victimization and delinquency found in the current study? One potential explanation is that bully victimization leads to delinquent behavior because victims feel powerless to control the actions of the bullies and want to feel powerful in other ways. Engaging in delinquent behavior such as stealing or vandalism may cause victims to feel less oppressed and more self-reliant because they are successfully defying the traditional rules of society. With respect to violent behavior such as assault, it may be that repeated bully victimization causes some children to reach a threshold at which point they fight back against their oppressors. The studies on school shootings by Vossekul et al. (2000) and Anderson et al. (2001) found that school shooters in the U.S. had frequently been targets of bullying by
their peers. Although the nature of the data in the aforementioned studies did not allow for this hypothesis to be tested, they suggest that prior victimization may have been a contributing factor to the violent outbursts. Yet another possible explanation for the current findings is that the relationship is indirect: bully victimization may lead directly to outcomes such as school disengagement and social withdrawal, which may be predictive of poor academic self-concept and achievement, truancy, lack of friendships, and emotional distress (e.g., Hodges & Perry, 1999; Wentzel & Caldwell, 1997). These outcomes may in turn lead to delinquent behaviors, thus acting as mediators or moderators of the relationship between bully victimization and delinquency.

Limitations and Future Research

A weakness of the current research is the possibility of imprecisely measured coefficients. In any observational study, omitted variable bias is a concern. Although 20 variables were incorporated into the estimation of the propensity scores, it remains true that we cannot adjust for unobservables, or unobserved risk predictors of bullying that were not measured by the NLSY97. In other words there may still be biases present, since it is only possible to detect (and subsequently control for) observable biases. This limitation is a concern with all observational studies. Ideally, other variables that would have been used in the creation of the propensity scores include measures of the parent-child relationship; assessments of self-esteem, impulsivity, submissiveness, and anxiety; school engagement; relationships with peers (such as size of the peer group, existence of a "best" friend, and measures of peer delinquency); peer ratings of popularity; and peer group perceptions of a youth's emotional and physical weakness. Using propensity score matching we can clearly state that the differences in delinquency outcomes between the bullied and non-bullied subjects are not the result of the set of covariates that were estimated. And to the extent that unobservables are correlated with the observable characteristics that were balanced, the unobservables are likely to be balanced as well.
Another limitation of the current study is the way in which the treatment variable was defined. A more precise definition of the "bully victimization" construct would be preferred. Although the NLSY97 definition does include use of the term "bullied" and the criterion for repeated incidents, a better measure would include a more detailed description of what is meant by victimization⁴⁹ or the use of multiple survey items to assess the construct on a scale. An ideal measure of the treatment variable would also have extracted more detailed information about the specific victimization that was experienced, for example, the number of times the bullying occurred, the number of different bullies that were involved, and the type of bullying that took place (e.g., physical or verbal, direct or indirect; see Chapter 1 of the dissertation for a typology of bullying behaviors). The location of the bullying would also be relevant, to ensure that respondents are indeed referring to bullying at school, as opposed to bullying in the home by siblings, or in afterschool programs, or on sports team, and so on. Furthermore, it would be useful to know exactly when the bullying experiences occurred. The survey item asked about bully victimization "prior to the age of 12 years," which precludes determining exactly how long before the baseline survey the victim suffered their bullying experiences. It may be that some individuals experienced bullying at age 8, while others were victimized at age 11, and still others may have been subjected to continued bullying from ages 8 to 11. It is possible that the timing of the victimization may impact the development of certain types of delinquent behavior. It may also be that some respondents were recalling incidents of victimization which occurred many years in the past, which raises the question of recall accuracy for experiences of bully victimization. Finally, another piece of relevant information is the degree to which victims also perpetrated bullying on others. The extent to which the NLSY97 victims were bullies themselves is unknown; the aforementioned research by Nansel and colleagues (2001) suggests that while 17 percent of students are bullied "sometimes" or more frequently during the school term, 6 percent of students are "bully-victims" and report both perpetration and victimization.

⁴⁹ For example, by prefacing the question with a definition of "bully victimization", along with examples of bullying behaviors. According to Olweus (1993c), true bullying requires repeated incidents along with intentional harm on the part of the bully, and a power differential between the bully and victim.
Although this study provides a rigorous analysis of a major U.S. dataset that includes recent information on bully victimization and delinquency among youths, additional evidence on the causal impact of victimization on delinquency is needed. Future work with the NLYS97 data could contribute by taking advantage of the panel data set and conducting a longitudinal analysis of the impact of early victimization on delinquency. In a more general sense, further research on bully victimization should involve longitudinal, prospective studies that employ more detailed definitions of bully victimization. Such designs would enable more precise estimations of treatment effects, as well as allow for sensitivity testing with regard to the impact of different types of bullying on the development of delinquent behaviors.

**Policy Implications**

The findings presented in the current study have important policy implications. They suggest that the experience of bully victimization in primary school is a contributing factor to the development of delinquent behavior in youths. As described previously, bully victimization accounts for between 2 percentage points to 8.2 percentage points of the behaviors in six of the delinquency outcomes studied. Table 27 presents approximations of the number of individuals who reported such behaviors who would not have engaged in the behavior had they not been victimized. For example, the analysis suggests that in the absence of bully victimization, 139 cases of theft, 106 cases of assault, and 41 cases of drug selling would have been avoided in this sample of respondents. Although these numbers may seem small, preventing 106 youths from committing assault would represent a reduction of nearly 17 percent of the 635 youths in the sample who reported engaging in such behavior. That this reduction would have positive implications for the youths who would have committed the assaults, along with their victims, other involved parties such as bystanders and parents, as well as society at large, is unquestionable.
Table 27 Delinquent Behavior Cases Attributable to Bully Victimization

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>% of treated engaged in outcome</th>
<th>% of outcome due to being bullied</th>
<th>Approximate # cases due to bully victimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Running away</td>
<td>19% (n=282)</td>
<td>3.2%</td>
<td>n= 50</td>
</tr>
<tr>
<td>Selling drugs</td>
<td>25% (n=418)</td>
<td>2.7%</td>
<td>n= 41</td>
</tr>
<tr>
<td>Vandalism</td>
<td>32% (n=544)</td>
<td>6.7%</td>
<td>n=116</td>
</tr>
<tr>
<td>Theft</td>
<td>38% (n=643)</td>
<td>8.2%</td>
<td>n=139</td>
</tr>
<tr>
<td>Other property crime</td>
<td>16% (n=272)</td>
<td>2.0%</td>
<td>n= 35</td>
</tr>
<tr>
<td>Assault</td>
<td>38% (n=635)</td>
<td>6.7%</td>
<td>n=106</td>
</tr>
</tbody>
</table>

In recent years schools, districts, counties, states, and countries have enacted anti-bullying legislation that may hold individuals and schools accountable for physical or emotional damage suffered as a result of uncontrolled school bullying. Thirty-five U.S. states have implemented anti-bullying legislation with respect to bullying in schools (Srabstein, Berkman, & Pyntikova, 2008). There are also numerous anti-bullying prevention and intervention programs that are currently being implemented at both the primary and secondary school levels, as well as a large number of programs targeting the development of prosocial skills and the prevention of deviant behavior.

Because the bulk of the research on victim impacts has focused on outcomes such as self-esteem, anxiety, depression, the development of minor physical ailments, school achievement, and peer bonding, intervention efforts for victims do not tend to focus on preventing the development of deviance. Because bullying is a type of aggressive behavior, intervention programs for perpetrators do frequently focus on strategies to reduce aggression and violence. The current study demonstrates that victims are not just at risk for internalizing problems, and so in a sense, victims should also be viewed as potential future offenders. As such, prevention and intervention efforts that seek to reduce adolescent delinquency should consider bully victimization a significant risk factor for offending. Prevention and intervention strategies that specifically target bullying would better serve by incorporating violence and antisocial behavior reduction strategies for both bullies and for victims.
Chapter 3
Are School-Based Bullying Prevention Programs Effective? A Meta-Analytic Review

INTRODUCTION

The Problem of School Bullying

Bullying among peers is remarkably common in schools worldwide and has been subject to a considerable amount of research over the past 15 years (e.g., Craig et al., 2000; Rigby, 1998; Salmivalli, Karhunen, & Lagerspetz, 1996a; Sharp & Smith, 1994). Along with physical aggression such as hitting, kicking, and hair-pulling, bullying includes non-physical aggression such as threatening, spreading rumors, and ostracizing, and in recent years has benefited from a proliferation in techniques associated with the popularity of online communication and mobile phones. In the United States, the most commonly cited study with respect to prevalence rates of bullying involves a nationally representative sample of over 15,600 students in grades 6-10. Results measured over the current school term indicate that 1 in 5 students have bullied others "sometimes" or more frequently, while almost 1 in 6 have been victims of bullying at least "sometimes" (Nansel et al., 2001).® These figures suggest that many millions of American students regularly participate in bullying or are victimized by their peers.

Bullying can have serious implications for both victims and perpetrators. Research suggests that these experiences may be related to a host of negative outcomes such as truancy, poor academic achievement, school drop-out, substance use, depression, anxiety, low self-esteem, eating disorders, injury, minor health ailments, suicide, delinquency, and violence (e.g., Card, 2003; Kaltiala-Heino, Rimpela, Marttunen, Rimpela, & Rantanen, 1999; Nishina et al., 2005; Wolke, Woods, Bloomfield, & Karstadt, 2000). Even before such serious ramifications

---

50 Prevalence rates of bullying and victimization vary widely across studies depending in part on research design, how bullying behaviors are defined, type of assessment(s) used to measure bullying and victimization, period of time over which bullying is examined, age of research participants, and other such considerations.
develop, daily fear of bullying is common in the lives of many school children. According to a study by the Kaiser Family Foundation (2001), youths aged 8 to 15 consider bullying a more serious problem in their own lives than are problems of discrimination, racism, and violence. Furthermore, the National Center for Education Statistics reports that 5 percent of students either skip school or avoid specific places in school because they are afraid (DeVoe et al., 2005), while Garrity et al. (1997) find that 20 percent of students are scared throughout much of the school day.

Bullying Prevention Efforts

Recently implemented anti-bullying policies range from school or district rules to state or country regulations, and such laws may hold individuals and schools accountable for physical or emotional injuries that students suffer as a result of school bullying. To date, 35 U.S. states have enacted anti-bullying legislation with respect to behavior in schools (Srabstein, Berkman, & Pyntikova, 2008). While laws vary by state, the majority require that schools establish rules prohibiting bullying on school premises and communicate such rules to students and parents, initiate mandatory reporting of incidents by teachers, develop an anonymous reporting system for students and parents, and implement a routine for investigating reports of bullying (Elinoff, 2004).

Along with federal, state, and local policies, bullying prevention efforts include whole-school interventions, classroom-based prevention programs, community-based anti-bullying programs, small-group interventions, and individualized approaches that target bullies or victims of bullying. School districts and school administrators seeking to reduce the level of bullying that takes place under their surveillance often choose to implement some type of school- or classroom-level bullying prevention policy or intervention. Such programs involve techniques such as increased monitoring of students, zero-tolerance bullying policies, peer mediation, counseling to at-risk students, and curriculum-based programs that provide general knowledge and impart skills designed to help students defend themselves against bullies.
(Juvonen, 2001). These anti-bullying prevention efforts differ considerably in the techniques used, the scope of the interventions, and the specific objectives they hope to achieve.

**Evidence-Based Prevention**

School-based bullying prevention efforts exhibit substantial range in their approach to the problem and make different assumptions about what works in preventing bullying. Despite their often disparate methods, these efforts are (presumably) sincere in their primary goal of reducing or eliminating bullying in schools (Juvonen, 2001). Do all such approaches work equally well? The answer is unequivocally no. Only a portion of the existing anti-bullying tactics have been evaluated with respect to their effectiveness, and of those evaluated, an even smaller fraction have been examined with a research design that possesses the requisite rigor for reaching cause and effect conclusions. The absence of evaluation may be due to many reasons; for example, lack of resources, lack of expertise, or pre-existing belief that the program is effective and does not require additional validation. The resistance or incapacity of schools to conduct proper program evaluations inevitably results in public and private funding of some anti-bullying efforts that are at best ineffective and a waste of student time, at worst exacerbating the problem, and either way indicative of a missed opportunity to effectively reduce bullying in schools. Evidence-based bullying prevention can only be implemented if the evidence both exists (i.e., the research is conducted) and the evidence provides a clear statement as to treatment effectiveness (i.e., the results of high quality research support the program’s benefits). Evaluation is imperative – tradition and experience are not a substitute for the results of scientific research (Farrington & Weisburd, 2007).

**The Current Study**

Of the rigorously evaluated bullying prevention programs, only a small number have been determined to have positive, significant impacts on the reduction of bullying or victimization. Results are not consistent from study to study, with some analyses reporting statistically significant and meaningful reductions in bullying behaviors as a result of an
intervention, and other studies reporting no noticeable change in rates of bullying or victimization. With such discrepant results, how can school-based bullying prevention programs be understood? Are they an effective approach to the problem, or a waste of valuable resources?

The current study seeks to determine the overall effectiveness of school-based programs for reducing bullying and victimization. Ideally suited for the task is the technique of meta-analysis, which produces a quantitative review of the cumulative research evidence. Through meta-analysis, the overall effectiveness of bullying prevention programs can be determined, and characteristics of programs that are linked with larger (or smaller) treatment impacts can be identified.

**Previous Reviews of Bullying Prevention Programs**

A preliminary literature search identified five studies involving systematic (using pre-determined selection criteria), narrative reviews of bullying prevention programs: Baldry and Farrington (2007), Rigby (2002a), Smith, Ananiadou, and Cowie (2003), Smith, Schneider, Smith, and Ananiadou (2004a), and Vreeman and Caroll (2007). In addition to these narrative reviews, two quantitative meta-analyses were identified: Ferguson, San Miguel, Kilburn Jr., and Sanchez (2007), and Ttofi and Farrington (2009). See Table 28 for an overview of these seven studies. First, Rigby's (2002a) study of bullying prevention efforts reviewed 13 rigorous studies conducted in North America, Western Europe, and Australia, and focused primarily on interventions involving young children. Programs were not limited by level of implementation, in other words, the review included programs implemented at the school, classroom, and individual level. Rigby concluded that with a few exceptions, the programs had a positive effect in reducing overall bullying behavior, with greater reductions found in rates of victimization rather than rates of bullying.

Smith et al. (2003) reviewed 10 evaluations of bullying prevention programs that involved more than one school and included pretest and posttest assessments of the bullying
and victimization measures. Across the 10 studies, the authors found mixed support for the effectiveness of bullying prevention programs.

The review by Smith et al. (2004a) limited the sample of studies to whole-school anti-bullying programs, and identified a set of 14 studies for inclusion in the meta-analysis. Analyses sought to determine whether program effectiveness was related to participant age, characteristics of the intervention, or characteristics of the research design. The authors reported that "effects fell almost exclusively into the categories of small, negligible, and negative" (p. 550). Also of note was that programs incorporating a monitoring component with respect to treatment integrity resulted in more positive effects than did programs without a monitoring component.

Baldry and Farrington's (2007) narrative review summarized what the authors deemed to be the "most important" evaluations of bullying prevention programs (16 studies), all of which used similar definitions of bullying, employed experimental or quasi-experimental research designs, and implemented both pretest and posttest measures or an age cohort design. The studies were required to include at least 200 students in the evaluation. Baldry and Farrington classified study outcomes into one of four categories, and found that 8 studies had desirable results, 2 had mixed results, 4 had small effects, and 2 had undesirable effects. They concluded that, overall, "the findings of existing evaluations of antibullying programs are optimistic rather than pessimistic" (p. 200).

The last of the narrative reviews, Vreeman and Carroll (2007), identified 26 studies for inclusion. Study outcomes were not limited solely to bullying and victimization; they also included aggressive behavior, violence, violent injuries, school responses to violence, student perceptions of school safety, and knowledge about or attitudes toward bullying. The authors reported fairly consistent evidence for the effectiveness of bullying prevention programs, with whole-school approaches and high levels of staff commitment playing important roles in the success of school-based interventions.

Two previous quantitative meta-analyses of bullying prevention program evaluations were identified. The analysis by Ferguson et al. (2007) included both anti-bullying programs as
well as more general violence-prevention programs. Only randomized experimental analyses published in peer-reviewed journals between 1995 and 2006 were eligible for inclusion, and only bullying outcomes were assessed (i.e., not victimization outcomes). These criteria resulted in a set of 23 studies that assessed nonviolent bullying and 25 studies that assessed outcomes of violence (some studies fell into both groups). Ferguson and colleagues reported that "although anti-bullying programs produce a small amount of positive change, it is likely that this change is too small to be practically significant or noticeable" (p. 408). The overall conclusion was that bullying prevention programs are not effective in reducing bullying or violence. The authors conducted additional analyses on moderator variables, and reported that treatment effect size was larger for at-risk students compared with low risk students (although still small), but was not associated with grade level of program implementation (elementary, middle or high school), or outcome measure (nonviolent bullying versus violent bullying).

Lastly, the study by Ttofi and Farrington (2009) included programs designed specifically for the prevention of bullying and/or victimization; evaluations were required to use self-report measures, a total sample size of at least 200 students, a control group, and an attempt at establishing internal validity through the use of randomization, pre-test measures, or an age-cohort design. A set of 30 studies were identified that met inclusion criteria, with 59 effect sizes extracted. The authors reported a significant overall pooled effect for both bullying and victimization, with a reduction of 17 to 23 percent in such behaviors for treatment group students. Ttofi and Farrington also identified several program components most strongly linked to a decrease in bullying and victimization, including parent training, playground supervision, anti-bullying rules in the classroom, peer mentoring or mediation, cooperative group work among professionals, disciplinary methods, classroom management, school conferences, videos, and information for parents.

51 The authors did not present a bibliography of the studies included in their analysis, so the number of studies that overlapped with the current analysis is unknown.
Table 28 Previous Narrative Reviews and Meta-Analyses of Bullying Prevention Programs

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study</th>
<th>Inclusion criteria</th>
<th># Studies</th>
<th># Studies overlap w/current analysis&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigby (2002a)</td>
<td>Narrative review</td>
<td>Focus on young children; pre- and posttest data or age cohort design</td>
<td>13</td>
<td>1</td>
<td>Generally positive but overall small effects</td>
</tr>
<tr>
<td>Smith (2003)</td>
<td>Narrative review</td>
<td>&gt;1 school; pretest and posttest measures</td>
<td>10</td>
<td>3</td>
<td>Mixed effects</td>
</tr>
<tr>
<td>Smith (2004a)</td>
<td>Narrative review</td>
<td>Whole-school programs</td>
<td>14</td>
<td>5</td>
<td>Small/ negative/ negligible effects</td>
</tr>
<tr>
<td>Baldry (2007)</td>
<td>Narrative review</td>
<td>200+ students; quasi-experimental or random assignment; pretest and posttest or age cohort design</td>
<td>16</td>
<td>5</td>
<td>Mixed, mostly positive effects</td>
</tr>
<tr>
<td>Vreeman (2007)</td>
<td>Narrative review</td>
<td>Outcomes not limited to bullying/victimization, also include aggressive behavior, violence, attitudes to and knowledge about bullying</td>
<td>26</td>
<td>9</td>
<td>Generally positive effects</td>
</tr>
<tr>
<td>Ferguson (2007)</td>
<td>Meta-analysis</td>
<td>Random assignment; antibullying and general violence prevention outcomes; bullying outcomes only</td>
<td>23</td>
<td>n/a; see footnote 51</td>
<td>Very small, non-significant effects</td>
</tr>
<tr>
<td>Ttofi (2009)</td>
<td>Meta-analysis</td>
<td>200+ students; self-reports; control group; use of pretest, randomization, or age-cohort design</td>
<td>30</td>
<td>15</td>
<td>Positive, significant effects for both bullying and victimization</td>
</tr>
</tbody>
</table>

<sup>a</sup> All studies reviewed in these seven reports were considered for inclusion in the current analysis. The lack of complete overlap in study sets is due to differences in the criteria used for study selection.

The current study continues the work of these comprehensive reviews of bullying prevention programs and expands the analysis in several different ways. Using meta-analysis, a quantitative summary of a systematically selected set of bullying prevention program evaluations is conducted. The approach allows for the determination of a precise, objective estimate of the treatment effect across the set of studies. This study differs from the meta-analysis by Ferguson et al. (2007) in that it allows for quasi-experimental research designs (as
well as experimental designs), includes studies from non-peer reviewed sources such as
doctoral dissertations, and focuses on bullying and victimization outcomes only (as opposed to
casting a wider net for outcomes of aggressive or violent behavior). The study differs from the
meta-analysis by Ttofi and Farrington (2009) by including studies with sample sizes as small as
20 per group and that were not limited to self-report measures of bullying and victimization.
Included studies were also required to have been published in the English language, and to
have used either randomization to groups or a pretest/posttest independent-groups design.

BACKGROUND ON META-ANALYSIS

What is Meta-Analysis?

First developed by Glass (1976; Smith & Glass, 1977), meta-analysis52 is a quantitative
literature review technique that uses explicit methods to identify relevant studies, then applies
objective methods to statistically analyze and synthesize study results. This approach differs
from a traditional narrative review, in which findings from multiple studies are discussed and
considered qualitatively, but are not combined in a numeric fashion. Fundamentally, the goal
of a meta-analysis is to limit bias and maximize objectivity in the identification and integration
of the collection of studies relevant to a specific research question, thereby improving the
generalizability of results and helping to resolve uncertainty when the results from primary
research or narrative reviews conflict.

Strengths of Meta-Analysis

When compared with narrative review, meta-analysis has several advantages. The first
is greater precision in estimating the effect in question. This is achieved through extracting and
converting results from individual studies into a common metric to enable cross-study
comparison, and quantitatively combining these results into a single effect estimate (or multiple

52 Also referred to as systematic review, research synthesis, research review, and quantitative review. The term is used more
narrowly by some writers, including this author, to refer to a quantitative analysis of a specific body of literature identified via
systematic review.
effect estimates separated into subgroups of studies). In comparison, narrative reviews often group the results of individual studies into themes and discuss them qualitatively, or engage in a pseudo-quantitative approach via some type of vote-counting – for example, counting the number of studies that report statistically significant versus null effects, or counting those that report positive versus negative findings. Clearly, this type of summation is preliminary, since it does not consider factors such as magnitude of the effects or differences in sample size. In general, using statistical significance to categorize studies that find effects versus those that do not is an unreliable technique (see Lipsey & Wilson, 2001). Whether a given result is deemed statistically significant depends on the size of the observed effect as well as the precision of the estimate. Since sampling error is a function of the size of the sample, due to low power studies with small samples are less likely to find statistically significant effects (Lipsey & Wilson, 2001). Meta-analysis circumvents this problem by ignoring individual study significance levels and focusing instead on the direction and size of the effects across studies. Because meta-analyses give greater weight to larger studies when estimating the combined effect across studies, the pooled effect estimate in a meta-analysis has considerably more statistical power than do the individual studies comprising the estimate (Lipsey & Wilson, 2001). Greater power translates into a greater likelihood of discovering meaningful treatment effects if they in fact exist.

Narrative reviews also face challenges in summarizing large or heterogeneous bodies of literature, for example, with respect to treatment evaluations in which variations in treatment style exist or differential treatment effectiveness is observed with different populations. Meta-analyses, on the other hand, easily summarize large amounts of information and include simple techniques for examining potential causes of variations in effect magnitude, such as causes attributable to study, treatment, measurement, or participant characteristics.

53 For example, a narrative review might vote-count by stating: "More than 75 percent of the studies reviewed reported positive findings for the bullying prevention programs, while fewer than 25 percent reported no effects. Overall, the literature supports such interventions as an effective way to reduce the problem of bullying in schools".

54 All else being equal; i.e., studies with small samples are less likely than studies with large samples to find statistically significant effects of the same size.
In addition, a properly-conducted meta-analysis is considerably more objective than a narrative review, since the technique requires the determination of explicit a priori criteria for study eligibility; in other words, the data collection is more systematic. A narrative review often does not implement inclusion criteria ahead of time, or may select studies by convenience, for example, studies the author is already familiar with, or has selectively chosen because they support a particular hypothesis. As such, it is hardly surprising that independent authors conducting narrative reviews of the same general topic often include a different set of studies and reach different summative conclusions when presenting their results (Judd, Smith & Kidder, 1991).

Finally, the explicit, transparent criteria set out for a given meta-analysis limits hidden biases and assumptions, and enables straightforward replication of the research by others (Lipsey & Wilson, 2001). In other words, armed with the identical inclusion/exclusion criteria and sources for study identification (such as the bibliographic databases to be searched), independent researchers could replicate the findings of the original study with relative ease. If two meta-analyses on the same topic produce different overall effect estimates or reach different conclusions, it should be possible to examine the study protocols to determine why they differed. Narrative reviews, on the other hand, are usually substantially more subjective in both study identification (e.g., they rarely use ex ante decision rules for inclusion) and interpretation (e.g., they may rely on vote-counting to categorize outcomes).

Weaknesses of Meta-Analysis

The most frequent criticism of meta-analysis is that analyses may pool together studies that are incommensurate – often dubbed "mixing apples with oranges" (Glass, 2000; Lipsey & Wilson, 2001). In other words, while possible, it is meaningless to produce a summary effect for studies that are conceptually dissimilar. A more detailed discussion of this issue is presented under Comparability of Study Outcomes, in the next section.

Another criticism of meta-analysis is the issue of publication bias. Also known as the "file drawer effect," this bias occurs when studies with statistically significant findings are
selected for publication more frequently than are studies with non-significant findings (Rosenthal & Rosnow, 1991). The result of this bias is that journals, government reports, and other peer-reviewed publications may present a skewed sample of the population of studies that have actually been conducted. To the extent that this is true the issue is of particular concern for meta-analysis, since the results of systematic searches of electronic databases and other such sources will fail to identify non-published, non-significant studies.

Meta-analysis has also been condemned for combining studies of differing methodological quality without accounting for the impact of research design on the observed outcomes (see Comparability with respect to research design in the next section). One approach to addressing this issue is to empirically test whether differences in study design impact the meta-analytic outcomes (see Other Data Coded). Another approach is to limit the meta-analysis inclusion criteria to those studies that meet a minimum level of methodological rigor in research design, such as studies that incorporate pretest measures, or studies that use random assignment to experimental groups. The drawback to requiring a high level of methodological rigor is that the number of studies eligible for inclusion in the analysis is likely to be reduced, sometimes substantially.

**Data Extraction**

**Comparability of Study Outcomes**

The raw data in a meta-analysis consist of the individual results from independent studies, which are abstracted in a standardized format. Since the analysis involves pooling data across studies, it is essential that the individual study outcomes be comparable with respect to conceptual nature, study design, and statistical form (Lipsey & Wilson, 2001).

**a) Comparability on a conceptual level**

First, the collection of studies in the meta-analysis must examine the same relationship. While the studies need not be pure replications of one another, the more alike they are the more compelling the case for claiming the comparison is of "apples to apples." However it is
common, especially in the social sciences, for studies to assess the same relationship in a broader sense – such as the effectiveness of general rehabilitative programs (which use a variety of approaches) on recidivism rates (Lipsey, 1999). Gene Glass, the grandfather of meta-analysis, surmised in 2000:

*The question of “sameness” is not an a priori question at all; apart from being a logical impossibility, it is an empirical question. For us, no two “studies” are the same. All studies differ and the only interesting questions to ask about them concern how they vary across the factors we conceive of as important. (para. 26)*

The initial step in ensuring a comparison of apples to apples is implementing specific inclusion and exclusion criteria during the selection of studies to incorporate in the meta-analysis. This process allows for the body of literature to be limited to studies that are comparable with one another with respect to characteristics such as method of treatment, location of intervention, age of participants, and so on. Even with the use of stringent selection criteria, some degree of heterogeneity among included studies should be expected. Including studies with a level of variety in the sense that they are not pseudo-replications of one another can be viewed as beneficial (Colditz, Burdick, & Mosteller, 1995). For one thing, finding consistency among heterogeneous studies indicates robustness of treatment effect. In addition, study differences can be examined from an exploratory perspective. With respect to evaluating the effectiveness of bullying prevention programs, this exploration may generate insights as to program elements or participant characteristics that appear important in producing larger treatment effects. Despite some potential benefits, too much between-study heterogeneity is undesirable and may lead to flawed conclusions if incompatible studies are pooled.
(b) Comparability with respect to research design

Second, it is generally not considered appropriate to pool findings from studies that use substantially different experimental designs to address the same basic research question. The concern is that studies that use different experimental designs may be estimating different population parameters (Morris & DeShon, 2002). For example, it would be inappropriate to pool findings from a study that used a randomized control design to assess the effects of a bullying prevention program with an observational study in which victimized students were interviewed in regard to their exposure to bullying prevention activities at school. While both of these studies would in principle assess the relationship between bullying prevention efforts and victimization, their differences with respect to research design would lead to expected disparities in the nature and interpretation of findings. In this case, the preferred method of analysis would be to limit study inclusion to one type of design only, or to conduct separate meta-analyses on sets of studies from each design type (see Lipsey & Wilson, 2001).

(c) Comparability with respect to statistical form: The effect size

Third, it is also not appropriate to combine study findings which appear in different statistical forms, e.g., beta coefficients, means, percentages, F-statistics, \( p \)-values, and so forth. For one thing, because test statistics and \( p \)-values are a function of sample size, they are not useful as standalone measures for the purpose of aggregation. Two studies demonstrating the same size of effect could produce quite different \( p \)-values based on differences in sample size. As discussed by Cohen (1988),

\[
I \text{ have elsewhere (1965) discussed the slippery slope of } "\text{If it’s statistically significant, it’s important, consequential, worth talking about, large, that is, significant!}\text{" Contributing to the low consciousness of effect size in large areas of behavioral science is the use of arbitrary measurement units. We rarely find ourselves dealing with dollars, years, centimeters, or bushels of manure. Another source of our difficulty is that, until recently, the standard output of many of our procedures has been a tau statistic, an } F, t, \text{ or } \chi
\]

55 Under such circumstances adjustments can be made, such as design-specific estimates of the sampling variance to account for differences in the precision of treatment effect estimates (Morris & DeShon, 2002).
square, together with the P values with (or without) their rewarding asterisks. A partial solution to this problem is the use of "pure" (dimensionless, unit-free) measures of effect size, what I called "rho" values in 1965. (p. 532)

For inclusion in a meta-analysis, all study findings are first transformed into the same format: the effect size. As pioneered by Jacob Cohen in 1962, an effect size is a standardized index that represents both the magnitude and direction of the outcome, in other words, how large of an effect the study found and whether the effect was positive or negative. Based on the concept of statistical standardization, each study produces an effect size statistic (or multiple effect sizes) derived from the particular quantitative outcomes presented in the research report. The derivation of each study’s effect size allows for a consistent interpretation of findings over the range of measures and variables employed across the set of included studies. This is similar in concept to standardizing test scores on a Z-distribution to allow for comparison with Z-scores from a different test of the same construct (Lipsey & Wilson, 2001).

Each effect size has a known distribution with an associated variance, important because characteristics of the sampling distribution (such as bias and sampling variance) are necessary to conduct a meta-analysis. The meta-analytic algorithm for the pooled effect size uses weighting; weights are determined by the inverse of the sampling variance. Sampling variance is a function of sample size, with larger studies contributing relatively more weight due to having smaller standard errors and thus providing more precise estimates about the effect of interest (Hedges & Olkin, 1985; Lipsey & Wilson, 2001). Standard errors are also necessary to estimate confidence intervals around the effect size statistic, as well as for other meta-analysis-related tests such as the Q-statistic for homogeneity (see the Analysis section for details on the Q-statistic).

As noted by Morris and DeShon (2002), effect sizes should all be scaled in the same metric, and even the use of a standardized effect size across studies does not ensure that scaling is comparable. For example, effect sizes can be computed using different standard deviations (e.g., the standard deviations of pretest scores versus posttest scores versus pooled
pretest/posttest scores), and effect sizes will only be comparable if the standard deviations are adjusted to a common metric.

**Effect size conventions**

Since the effect size represents the magnitude of the treatment effect determined by a study, it is natural to question whether a given effect size is small or large.\(^{56}\) The well-recognized gospel, according to Cohen, is that a small effect size is roughly 0.20 standard deviations, medium is .50 standard deviations, and large is .80 standard deviations or above (Cohen, 1988). Despite near universal acceptance, Cohen himself has stated "The values chosen had no more reliable a basis than my own intuition. They were offered as conventions because they were needed in a research climate characterized by a neglect of attention to issues of magnitude" (Cohen, 1988, p. 532). Cohen's conventions appear to be supported in the literature; empirical work by Mark Lipsey (1990) suggests a very similar range, with small effect sizes defined as .15 standard deviations or below, medium at .45 standard deviations, and large effect sizes as .90 standard deviations or above. Lipsey developed these categorizations based on the distribution of average effect size estimates from 102 meta-analyses involving 6,700 individual studies and almost 800,000 subjects.

**Effect Size Calculations**

Several types of effect size measures exist, and each typically has more than one method of computation. Commonly used effect size measures include the standardized mean difference (such as Cohen's \(d\) or Hedges' \(g\)), the odds-ratio, the correlation coefficient, the proportion difference, and the relative risk. The choice of effect size measure depends on both the research question of interest, as well as the data reported in the study. Extracting the relevant data from primary research reports to compute effect sizes may prove challenging, as

---

\(^{56}\) Bloom, Hill, Black & Lipsey (2008), in an assessment of effect sizes in educational interventions, argue that heuristics such as Cohen's standards are inappropriate, "Their one size fits all character is not sufficiently differentiated to be useful for any specific intervention circumstance and is more likely to result in misleading expectations and interpretations about the respective effect size" (p. 30). Bloom et al. contend that different interventions have different goals, and these goals should be considered when interpreting the practical significance of effect sizes that result from program evaluations. An effect size of .50 might be considered "medium" by Cohen's standards, but in practice might reflect a finding in which a school's annual achievement growth was doubled.
authors rarely present identical types of information and may not provide the requisite data to
derive effect size estimates. When a key piece of data is omitted from a research report, it
implies that either the information was not collected in the first place, or was collected but was
not reported by the authors. There are three options for dealing with this scenario: (1) exclude
the study from the meta-analysis due to lack of data; (2) attempt to contact the author(s) to
retrieve the necessary data; or (3) estimate the data from other information in the report.
Option (1) is the most conservative, following the argument that the meta-analysis protocol is
developed a priori and should be strictly followed to avoid bias in study selection. Yet this
approach is still open to bias, since different journals have different reporting guidelines;
furthermore, publication types other than journals (e.g., government reports) may diverge in
the types of information they deem of interest to the reader. As such, excluding studies that
otherwise meet all inclusion criteria but fail to report one or more pieces of information
necessary to compute the effect size or standard error may not be the best option. Option (2)
is ideal in theory yet is often impractical; it may be difficult to get in touch with authors and
obtain meaningful contributions from them in regard to study details, particularly if the study
was published several years in the past. In light of these limitations Option (3) is commonly
used in meta-analyses (e.g., Borman, Hewes, Overman, & Brown, 2002); whereby certain
assumptions about omitted pieces of information are made in order to generate the required
data.

The following sections describe different methods to calculate effect sizes and their
associated standard errors, focusing on the effect size measures used in the current study.

**Continuous outcome measures: The standardized mean difference**

The standardized mean difference effect size applies to studies that compare two
groups (e.g., treatment and control) on their respective mean scores on an inherently
continuous measure that is assessed in different ways across the set of study samples.\(^57\) The

---

\(^{57}\) For a set of studies that operationalizes the dependent variable in the same way, the unstandardized mean difference effect
size can be used. This measure was not applicable in the current analysis, due to variation in the outcome measures used to
assess bullying and victimization.
most commonly used standardized mean difference measure is Cohen's $d$, in which the effect size is computed as the mean of the treatment group minus the mean of the control group divided by the pooled standard deviation:

\[
Cohen's\ d = \frac{\bar{Y}_T - \bar{Y}_C}{\sigma_{pooled}},
\]

where $\bar{Y}_T$ is the mean score of the treatment group, $\bar{Y}_C$ is the mean score of the control group, and $\sigma_{pooled}$ is the pooled population standard deviation (Cohen, 1988).

A closely related measure of the standardized mean difference is Hedges' $g$ (Hedges, 1981, 1982b), which takes into account the upward bias in effect size that Hedges demonstrated as a concern when effect size is derived from small sample sizes (due to fewer degrees of freedom used in the estimation). Hedges provides a correction for this bias as follows:

\[
Hedges'\ g = \left[1 - \frac{3}{4n - 9}\right] \left(\frac{\bar{Y}_T - \bar{Y}_C}{SD_{pooled}}\right)
\]

where $n = n_T + n_C$, $n_T$ and $n_C$ are sample sizes for the treatment and control groups, respectively, and $SD_{pooled}$ is the pooled estimate of the standard deviation of the treatment and control groups, defined by:

\[
SD_{pooled} = \sqrt{\frac{(n_T - 1)SD_T^2 + (n_C - 1)SD_C^2}{n_T + n_C - 2}},
\]

58 A third commonly used measure is Glass's $\Delta$, defined as $[(\bar{Y}_T - \bar{Y}_C)/S_{control}]$, where $S_{control}$ is the standard deviation of the control group (Glass, McGaw, & Smith, 1981).

59 The pooled standard deviations in the denominators of Cohen's $d$ and Hedges' $g$ are very similar; the difference is whether the population estimate or the (inferential) within-sample estimate is used (Rosnow & Rosenthal, 2003).
in which and $SD_T^2$ and $SD_C^2$ are standard deviations for the treatment and control groups.

In the current analysis, the standardized mean difference effect size was calculated, but with an adaptation fit for the type of studies included – the pretest-posttest control group design. This formula takes advantage of the baseline data available in the included studies, which may allow for a less biased estimate of the treatment effect size if there are initial pretest group differences. In other words, the estimate assesses the difference between the standardized mean changes from pretest to posttest in the treatment and control groups. The standardized mean difference effect size is calculated as follows:

$$ES_{SMD} = c_p \left[ \frac{(\bar{Y}_{T,\text{post}} - \bar{Y}_{T,\text{pre}}) - (\bar{Y}_{C,\text{post}} - \bar{Y}_{C,\text{pre}})}{SD_{\text{post}}} \right],$$

where $\bar{Y}_{T,\text{post}}$ and $\bar{Y}_{C,\text{post}}$ are the mean scores at posttest for the treatment and control groups, respectively, $\bar{Y}_{T,\text{pre}}$ and $\bar{Y}_{C,\text{pre}}$ are the mean scores at pretest for the treatment and control groups, and $SD_{\text{post}}$ is the pooled posttest standard deviation for the treatment and control groups.

The pooled standard deviation is defined as follows:

$$SD_{\text{post}} = \sqrt{\frac{(n_T - 1)SD_{T,\text{post}}^2 + (n_C - 1)SD_{C,\text{post}}^2}{n_T + n_C - 2}},$$

where $SD_{T,\text{post}}^2$ and $SD_{C,\text{post}}^2$ are the squared standard deviations for the treatment group and control group at posttest, respectively.
The pooled bias correction, $c_p$, is defined as:

$$c_p = 1 - \frac{3}{4(n_T + n_C - 2) - 1}.$$ 

The standard error of $ES_{SMD}$ is:

$$SE_{SMD} = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES)^2}{2(n_T + n_C)}}. \quad (60)$$

**Standardized mean difference conversion: For continuous outcomes studies that do not report means and standard deviations**

For studies that do not present means and standard deviations for the treatment and control groups at pretest and posttest, a conversion formula may be used to compute the effect size from available test statistics (Lipsey & Wilson, 2001). In the current analysis one study provided product moment correlations between the bullying outcome variable and the treatment/control group variable (Menard, Grotpeter, Gianola, & O’Neal, 2008). With this data, the effect size can be computed as follows:

$$ES_Y = \left[\frac{r_{post}}{\sqrt{1 - r_{post}^2}}\right] - \left[\frac{r_{pre}}{\sqrt{1 - r_{pre}^2}}\right],$$

where $r_{post}$ is the product moment correlation at posttest, $r_{pre}$ is the product moment correlation at pretest, and $p$ is the proportion of subjects in the treatment group.

\(60\) This standard error formulation ignores the sampling error implications of the pretest adjustment, but is a reasonable approximation of the unbiased standard error for $ES_{SMD}$ (M. Lipsey, personal communication, February 21, 2009).
The standard error is calculated using the same formula as for $ES_{simp}$:

$$SE_r = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES)^2}{2(n_T + n_C)}}.$$  

**Dichotomous outcome measures: The odds ratio with Cox transformation**

The second type of study included in the current meta-analysis uses dichotomous outcome measures; authors typically present raw numbers or percentages of treatment and control group subjects who were bullied and/or victimized at pretest and posttest. An effect size for this type of data can be calculated as an odds ratio, which compares the difference between proportions of successful (i.e. non-bullied or non-victimized) outcomes associated with each group. The odds ratio is defined as:

$$OR = \frac{p_T(1 - p_C)}{p_C(1 - p_T)},$$

where $p_T$ is the proportion of the treatment group that was not bullied (or victimized) and $p_C$ is the proportion of the control group that was not bullied/victimized.

A small obstacle is that the odds ratio measure of effect size is not numerically compatible with the standardized mean difference (Lipsey & Wilson, 2001). One solution to this problem, as long as the outcome is assumed to be inherently continuous, is to apply a transformation to the bivariate dependent variable that adjusts for the dichotomization. Another common approach would be to artificially dichotomize the continuous outcome variables to make them comparable with the dichotomous variables.61
Moscoso, 2003). In the current analysis, the Cox transformation was used, in which the log odds ratio\(^{62}\) \((L_{or})\) is divided by the constant 1.65 (see Cox, 1970, and Sanchez-Meca et al., 2003). After applying the adjustment for pretest differences, the Cox-transformed effect size becomes:

\[
ES_{Cox} = \left[ \frac{L_{OR,post}}{1.65} \right] - \left[ \frac{L_{OR,pre}}{1.65} \right],
\]

where \(L_{OR,post}\) is the log odds ratio at posttest and \(L_{OR,pre}\) is the log odds ratio at pretest.

The standard error for the Cox-transformed odds ratio is:

\[
SE_{Cox} = \sqrt{0.367 \left[ \frac{1}{O_{1T}} + \frac{1}{O_{2T}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right]},
\]

where \(O_{1T}\) and \(O_{1C}\) are the successful outcome frequencies (i.e., non-bullied or non-victimized) in the treatment and control groups, respectively, and \(O_{2T}\) and \(O_{2C}\) are the non-successful outcome frequencies (i.e., bullied or victimized) in the treatment and control groups.

The Cox transformation produces an effect size for the difference between proportions that is comparable to the standardized mean difference effect size; the transformation adjusts the effect size to be independent of how \(p\) and \(1-p\) are located between

---

\(^{62}\) More accurately, the natural logarithm of the odds ratio.
0 and 1. After applying this transformation, the effect sizes can all be pooled and analyzed together.\textsuperscript{63}

**Extracting Other Data**

A controversial issue in meta-analysis is whether to consider, and if so, how to consider, the methodological quality of the set of included studies. In a typical primary study, if the raw data are flawed then the conclusions drawn from any analyses of the data are likely to be faulty as well. Presumably, then, if the raw data for a meta-analysis are flawed – the raw data being the outcomes from the set of included studies – then the conclusions drawn from the meta-analysis may be biased (Juni, Altman, & Egger, 2001). One way to approach the issue of methodological quality is to formally weight each study by rigor, as assessed on some sort of quality scale. Then, studies rated as "high quality" would be given greater weight in the pooled effect size analysis, while studies rated as "low quality" would contribute less weight. Although this approach may sound tempting, quality is a complex issue that is not straightforward to define. Quality may refer to aspects of study design (e.g., issues of internal validity such as selection or maturation bias, Cook and Campbell, 1979), treatment implementation fidelity, data collection, data analysis, or reporting of results. There are numerous existing composite quality rating scales that combine various study elements into a single numerical score, such as the Maryland Scale of Scientific Methods (Sherman et al., 1998). These scales vary substantially in terms of the dimensions they address and in their level of detail and complexity; there exists no universal agreement as to what elements – let alone what weighting of elements – are sufficient for an overall rating of a study's "quality" (Juni et al., 2001). For these reasons, many meta-analysts (e.g., Greenland, 1994; Juni et al., 2001; Lipsey & Wilson, 2001) do not recommend using or weighting studies by composite quality measures.

An alternative approach to addressing quality is to code a selection of study elements deemed important and potentially influential for the individual effect size estimates, and then

\textsuperscript{63} Lipsey and Wilson (2001) suggest the inclusion of a dummy variable in meta-regression models to test for whether the method of effect size estimation results in differences in the mean effect size estimates. If differences are found, it may be prudent to split the continuous and dichotomous dependent variables into subgroups and analyze them separately.
conduct separate meta-analyses on subgroups of studies, such as those using random assignment versus non-random assignment to experimental groups, studies with large sample sizes versus those with small sample sizes, and so on. This approach was taken in the current study and involved coding key pieces of information for each effect size, such as specifics regarding subject assignment to treatment and control groups, the type of measure used to assess program outcomes, and whether or not the study was published in a peer-reviewed source.

Data Analysis

Pooled Analysis

The meta-analysis algorithm weights the set of transformed effect sizes by their inverse variance weights. More specifically, each study has an effect size (ES) which has an associated standard error (SE) that represents the precision of the effect size. The standard error is used to create the inverse variance weight, \( w \); weighting is used so that more precise effect size estimates contribute more to the pooled mean effect than do less precise effect sizes.\(^6^4\) Since larger sample sizes are associated with smaller variances, the treatment effect estimates from large samples are more precise. It is therefore advantageous for these estimates to contribute more to the pooled mean effect. The inverse variance weight is calculated with the following equation:

\[
w_i = \frac{1}{SE_{i}^2} = \frac{2n_{T,j}n_{C,j}(n_{T,j} + n_{C,j})}{2(n_{T,j}n_{C,j})^2 + n_{T,j}n_{C,j}(ES_i)^2},
\]

where the subscript \( i \) refers to individual effect size estimates.

\(^6^4\) An effect size derived from a sample of 1,500 subjects would be considered a more precise estimate of the population parameter than an effect size derived from a sample of 15 subjects (ceteris paribus).
The meta-analysis mean effect size across all studies is calculated by:

\[
\bar{ES} = \frac{\sum_{i}^{k} (w_i \cdot ES_i)}{\sum_{i}^{k} w_i},
\]

in which each study's effect size is multiplied by its inverse variance weight \((w_i \cdot ES_i)\), summed across all studies \(i\) through \(k\), and divided by the sum of the inverse variance weights for studies \(i\) through \(k\).

The normal distribution is used to test the significance of the mean effect size; \(Z\) is defined as follows:

\[
Z = \frac{\bar{ES}}{SE_{\bar{ES}}},
\]

where the standard error of the mean effect size is defined by:

\[
SE_{\bar{ES}} = \sqrt{\frac{1}{\sum_{i}^{k} w_i}}.
\]

---

65 In the fixed effects model, as discussed on p. 209.
The upper and lower bound 95% confidence intervals, which indicate the precision of the mean effect size by designating the range within which the population mean is likely to fall in the observed data, are:

$$95\% \text{ C.I.} = \overline{ES} \pm 1.96(SE_{ES}) .$$

If the confidence interval for the mean effect size does not cross zero, then we reject the null hypothesis of no treatment effect.

**The Homogeneity Statistic (Q)**

Because the effect sizes used in a meta-analysis are pooled into a mean treatment effect across all studies, it is important to determine whether these effect sizes all estimate the same population effect (Lipsey & Wilson, 2001). This determination can be achieved through a test for homogeneity of the effect size distribution, using the $Q$-statistic. In essence, this statistic tests whether it is reasonable to assume that all of the effect sizes in the meta-analysis are estimating the same population parameter. $Q$ is defined as follows:

$$Q = \sum_{i} w_i \left( ES_i - \overline{ES} \right)^2 \sim \chi^2_{(k-1)} ,$$

where $ES_i$ is each individual study's effect size, $w_i$ is the inverse variance weight for $ES_i$, and $\overline{ES}$ is the weighted mean effect size across all studies. $Q$ is distributed as a chi-square with $k-1$ degrees of freedom, where $k$ is the number of effect sizes.

Essentially, $Q$ is just a weighted sum of squares; the null hypothesis being that all of the individual study effect sizes are estimating the same underlying population parameter. If the observed $Q$ is less than the critical value of the $Q$-statistic, we fail to reject the null hypothesis of homogeneity. The logic is as follows: If the studies are measuring the same effect and the
distribution is homogeneous, then random subject-level sampling error should be the only reason for the differences between individual effect sizes (i.e., because studies drew different samples of participants). In this case, $Q$ will not be statistically significant.

On the other hand, $Q$ will be statistically significant if the effect sizes vary more than would be expected due to sampling error, which suggests that the studies are not estimating the same underlying parameter.\textsuperscript{66, 67} In this case, an analysis that calculates an overall mean effect size pooled across studies is not a good approach to summarizing the distribution. In the case of heterogeneity, the primary options to approach the problem are to fit a random effects model, model the between study differences, or both. These options are discussed below.

**Fixed Effects versus Random Effects Models**

As previously discussed, weighting each study's effect size by its inverse variance is a procedure designed to emphasize study precision. There are two main meta-analysis models available to accomplish this goal: "fixed effects" and "random effects". The essential difference in these models is the manner in which the variability of results among studies is handled (Egger & Smith, 2001). If the $Q$-statistic is non-significant, and the studies are similar to one another with respect to treatment, sample characteristics, and so on, a fixed effects model is usually recommended. In this model, the assumption is that all of the studies are estimating the same population parameter (the effects are "fixed"), and any between-study variability is a result of random variation, or "noise" (Deeks, Altman, & Bradburn, 2001). When calculating the mean effect size, each study is weighted by its relative precision (inverse variance weight), as described earlier, with:

\[ I^2 = \frac{Q - (k - 1)}{Q} \times 100 \]

\textsuperscript{66} Meta-analyses sometimes assess the magnitude of heterogeneity of the included studies, as calculated by the $I^2$ statistic (Higgins, Thompson, Deeks & Altman, 2003); $I^2 = \frac{Q - (k - 1)}{Q} \times 100$. By convention, all negative values of $I^2$ are recoded as 0. According to Higgins et al. (2003), $I^2$=25\% represents a small degree of heterogeneity, $I^2$=50\% is moderate heterogeneity, and $I^2$=75\% is a large degree of heterogeneity.

\textsuperscript{67} The $Q$ test for homogeneity can be subject to low statistical power (Higgins et al., 2003), or even "too much" power. For example, $Q$ can be non-significant with a large $F$, which suggests low power, or significant with a small $F$, which suggests "too much" power (Valentine, 2008).
In contrast, the random effects model (DerSimonian & Laird, 1986) is appropriate if the set of studies is reasonably heterogeneous and the $Q$-statistic for homogeneity is significant, indicating that the observed heterogeneity among studies is more than would be expected due to random subject-level sampling error. This model assumes that the variability among individual study effect sizes is due to sampling error across studies (noise), as well as variability in the population effects being estimated – in other words, the studies are estimating different population parameters (the effects are "random"; Egger & Smith, 2001).

In a random effects model, the inverse variance weights take into account both within-study variability (subject-level sampling error in estimating the population parameter; the fixed effects variance), plus between-study variability (differences in the actual parameters being estimated; the random effects variance; Sweet & Applebaum, 2004). The random effects model calculates weights as the inverse of the sampling variance plus a constant ($v$) that represents the between-study variability in population effects:

$$w_{RE,i} = \frac{1}{SE_i^2 + v}$$

---

68 If the assumptions of the fixed effects model are exactly met, estimates from the random effects model will be equivalent to estimates from the fixed effects model (Colditz, Burdick, & Mosteller, 1995; Lipsey & Wilson, 2000)
The random effects variance component, $v$, is determined as follows:\footnote{This formula is a noniterative method based on the method of moments; an alternative method for estimating the random effects variance component uses maximum likelihood (Lipsey & Wilson, 2001).}

\[
\hat{v} = \frac{Q - (k - 1)}{\sum_{i=1}^{k} w_i - \left( \frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i} \right)},
\]

where $k$ is the number of effect sizes.

The DerSimonian and Laird random effects model results in smaller studies being weighted more heavily than they are in the fixed effects model, because small studies highlight the distribution of effects across the set of studies (Higgins & Green, 2006). In other words, the manner in which the model assigns study weights is responsive to the amount of heterogeneity observed between studies, with an increased level of heterogeneity resulting in a greater number of near-equivalent weights (Colditz et al., 1995). In practice, the random effects model is more conservative than the fixed effects model and generally has lower statistical power with wider confidence intervals, since the between-study variability that is added to the sampling error variability increases the uncertainty associated with the mean estimate (Lipsey & Wilson, 2001). Consequently, the random effects model tends to give fewer statistically significant results.

**Modeling Between-Study Differences**

As discussed previously, statistical heterogeneity refers to the situation in which the true effects estimated by each study in a meta-analysis are not identical (Thompson & Higgins, 2002). Heterogeneity can be approached by implementing a random effects model (described above), modeling between-study differences, or both. The modeling of study differences is
helpful if there is reason to believe that one or more study-level variables may "explain" part of the between-study heterogeneity, in other words, if the observed heterogeneity is beyond what would be expected due to random subject-level sampling error alone. In this situation, a fixed effects model can be implemented in conjunction with the assumption that the variability is systematic, and the sources of this variation can then be explored (Lipsey & Wilson, 2001).\textsuperscript{70,71}

The two major techniques for modeling between-study differences are subgroup analyses (also called stratified analysis; Deeks et al., 2001) and meta-regression models. Either or both of these methods may be used to identify sources of the variability in results across the set of effect sizes, which enhances the results of a meta-analysis through exploring factors that predict treatment success and failure.

**Subgroup analyses**

With a categorical indicator that exhibits between-study variation,\textsuperscript{72} e.g., randomized versus non-randomized research design, full-year versus half-year program, or primary versus secondary school students, the studies may be grouped according to the particular variable and a separate meta-analysis conducted on the results within each subgroup (Deeks et al., 2001).\textsuperscript{73}

Each subgroup analysis produces a pooled treatment effect estimate, and these estimates can be compared to determine whether the grouping variable is related to differences in pooled effect estimates. If there is evidence for variation in the treatment effect between subgroups, for example, the subgroup of studies assessing effects on males results in a substantially larger treatment effect than the subgroup of studies assessing effects on females, this suggests that

---

\textsuperscript{70} Some authors contend that the $Q$-test for homogeneity has low power and so a non-significant $Q$ is not necessarily an indication that between-study variables should not be examined (e.g., Higgins et al., 2003).

\textsuperscript{71} Another option is to implement a mixed effects model, which assumes that there is variability in study effect sizes beyond random sampling error, and that this variability stems from both systematic and random sources. This model tests whether selected study or sample characteristics act systematically as moderator variables in the derivation of effect sizes, while simultaneously allowing for an element of random residual variance to remain. The residual variance component is estimated and then incorporated into the inverse variance weight for each study.

\textsuperscript{72} The variable can be related to characteristics of the study, population, or treatment.

\textsuperscript{73} The grouping variables should be defined ex ante, based on theoretical relevance, as opposed to ex post data searching for variables that appear to be potential sources of between-study differences (Colditz, Burdick, & Mosteller, 1995). Pre-specification of variables will decrease the chances of a type I error, which is more likely when multiple groupings of variables are being compared without theoretical or empirical justification.
gender is an important source of between-study heterogeneity and may have a moderating effect on treatment efficacy (Deeks et al., 2001).

The separate meta-analytic estimates in the subgroups should only be compared informally, by assessing the magnitudes of the pooled effect sizes. It is incorrect to conclude that a subgroup factor explains heterogeneity if statistical significance is found for either the pooled effect or the $Q$-test in one group but not for the other group (Higgins & Green, 2006). The subgroups are unlikely to be equivalent in the amount of information they are providing, and consequently they will not have equivalent power with which to detect potential treatment effects. An appropriate assessment of the influence of a grouping variable involves formally comparing the groups with one another using a $Z$-statistic (Deeks et al., 2001):

$$Z = \frac{ES_1 - ES_2}{\sqrt{(SE(ES_1))^2 + (SE(ES_2))^2}}$$

where $ES_1$ and $ES_2$ are the estimates of the overall treatment effect within each subgroup, and $SE(ES_1)$ and $SE(ES_2)$ are the standard errors of these estimates.

Another method by which to compare subgroups is analogous to an ANOVA (Hedges, 1982a). As described by Lipsey and Wilson (2001), using this method the total homogeneity statistic, $Q_T$ (defined earlier as $Q$), is partitioned into two segments: the segment that can be explained by the between-groups moderator variable, $Q_b$, and the pooled within groups segment, $Q_w$.

$Q_b$ is a weighted sum-of-squares of the mean effect size for each subgroup of the moderator variable around the pooled mean, and is calculated as follows:
\[ Q_B = \left( \sum w_j ES_j^2 \right) - \left( \frac{\sum w_j ES_j}{\sum w_j} \right)^2, \]

where \( ES_j \) is the weighted mean effect size for each subgroup, \( \Sigma w_j \) is the sum of the inverse variance weights within each subgroup, and \( j = 1, 2, 3, \ldots \) up to X number of groups defining the categorical variable (Lipsey & Wilson, 2001).

In contrast \( Q_W \) is a weighted sum-of-squares of the individual study effect sizes within each subgroup around the subgroup mean, pooled over the \( j \) groups. First the mean effect size, standard error, and \( Q \) statistic are estimated for each subgroup of effect sizes, as follows:

\[ Q_{\text{group1}} = \sum w_i (ES_i - ES_j)^2, \]

where \( ES_i \) is the individual study effect size, \( ES_j \) is the weighted mean effect size for each subgroup, \( w_i \) is the weight for each effect size, \( i \) equals 1, 2, 3, \ldots up to X number of effect sizes, and \( j \) equals 1, 2, 3, \ldots up to X number of groups (Lipsey & Wilson, 2001).

The separate \( Qs \) for each subgroup of effect sizes are then summed to form \( Q_W \):

\[ Q_W = Q_{\text{group1}} + Q_{\text{group2}} \sim \chi^2_{k-j}, \]

distributed as a chi-square distribution, where \( k \) is the number of effect sizes and \( j \) is the number of groups.
$Q_B$ can also be found by using basic subtraction; $Q_B$ is simply the difference between $Q$-total ($Q_T$) and $Q_W$:

$$Q_B = Q_T - Q_W \sim \chi^2_{j-1}.$$  

In other words the total heterogeneity is partitioned into the portion determined by differences between the subgroups, and the portion attributable to differences within the subgroups. These estimates are then tested. If the between category variance ($Q_B$) is statistically significant, it suggests that the mean effect sizes produced by each subgroup are significantly different from one another – and differ by more than would be expected from sampling error. If $Q_W$ is not statistically significant, it suggests that subject level sampling error accounts for the remaining variability among effect sizes, which indicates that the categorical variable captured in $Q_B$ is responsible for explaining the excess heterogeneity in the distribution of effect sizes.

**Meta-regression**

In the case of continuous variables or multiple between-study categorical variables to be analyzed simultaneously, a modified weighted least squares regression analysis may be used to model between-study differences (Hedges & Olkin, 1985). As in standard linear regression analysis, the coefficients derived in a meta-regression describe how the dependent variable changes with a unit increase in the predictor variables. The statistical significance of each regression coefficient represents a test of whether a linear relationship exists between the predictor variable and the outcome.

Similar to meta-analysis, in meta-regression studies are weighted by the precision of their effect estimates; thus, larger studies have more influence on the dependent variable than do smaller studies (Higgins & Green, 2006). In a random effects meta-regression, also called a

---

74 There should be a sufficient number of studies used in a meta-regression model, with some authors recommending at least ten studies per group (Higgins & Green, 2006).

75 Dummy variables can also be used in a meta-regression to investigate categorical variables that identify multiple subgroups (Higgins & Green, 2006).
"mixed model," two additive components of variance are included in the model. Along with a within-study variance term, a random effects term is included which allows for residual heterogeneity – i.e., variation in the prediction of the treatment effect that is not explained by the predictive covariates (Sutton, Abrams, Jones, Sheldon, & Song, 2000). This is similar to the random effects variance component in a random effects meta-analysis model.

The random effects meta-regression model is described following the outline put forth by Sutton and colleagues (2000). Suppose the set of studies includes \( k \) independent effect size estimates \( ES_1, \ldots, ES_k \). Each estimate \( ES_i \) is defined by \( ES_i = \theta_i + \varepsilon_i \) where \( \theta_i \) is the true effect size for each of the \( k \) studies, \( i=1, \ldots, k \), and \( \varepsilon_i \) is the error term (the \( \varepsilon_i \) are assumed to be independent, each with \( \mu=0 \) and variance \( \sigma^2 \)). The variance for each of the \( ES_i \) effect size estimates is \( \text{Var}(T_i) = \tau^2 + \nu_i \), where \( \tau^2 \) is the between-study (random effects) variance and \( \nu_i \) is the within-study variance. The weighted least squares algorithm calculates weights as the reciprocal of each study's variance:

\[
W_i = \frac{1}{\nu_i} = \frac{1}{(\tau^2 + \nu_i)}. 
\]

The linear model is outlined as follows:

\[
\theta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \mu_i,
\]

where \( X_{i1}, \ldots, X_{ip} \) are the values of the \( p \) predictor variables \( X_1, \ldots, X_p \) for the \( i \)th study, \( \beta_0 \) is the model intercept, \( \beta_1, \ldots, \beta_p \) are regression coefficients for the predictor variables \( X_1, \ldots, X_p \) and \( \mu_i \) is the random effect of each individual study, representing the deviation of the study's true effect from the value predicted by the model. Each random effect \( \mu_i \) is assumed to be independent and normally distributed, with variance \( \sigma^2_\mu \).
To estimate the parameters for each effect size $ES_i$, we substitute:

$$ES_i = \hat{\beta}_0 + \hat{\beta}_1X_{i1} + \hat{\beta}_2X_{i2} + \ldots + \hat{\beta}_pX_{ip} + \mu_i + \epsilon_i$$

This equation has two components in its error term, $\mu_i + \epsilon_i$, representing the residual in the individual $ES_i$ estimation. See Sutton et al. (2000) for details on estimating the $\hat{\beta}$s and the random-effects variance ($\tau^2$).

**Sensitivity Analysis**

As the last step in a meta-analysis, the robustness of the meta-analytic findings is assessed. Two popular options include an examination of publication bias and an assessment of solution sensitivity.

**Publication bias**

It is well-known that only a portion of completed research is published in peer-reviewed journals or other publications that are indexed in electronic databases. In particular, studies with statistically significant results are more likely to be published, which suggests that meta-analyses relying on bibliographic sources to identify candidate studies may end up producing a biased treatment effect (Egger, Dickersin, & Smith, 2001). Related to this is the finding that small-scale studies tend to show larger treatment effects than do large-scale studies (also known as "small study effects"). The reasons for this are not clearly known and tend to differ across different intervention and research contexts, sometimes due to factors such as the studies targeting specific high-risk groups for which treatment is more likely to have an impact (Sterne, Gavaghan, & Egger, 2000). The likelihood of a smaller study being published is greater if the study rejects the null hypothesis and reports a stronger treatment effect, and this may contribute to the problem of publication bias.

In a meta-analysis, the possibility of both publication bias and small study effects can be assessed through a funnel plot, in which each study's effect size is plotted against its standard error. This plot allows for a visual inspection of the relationship between treatment
effect size and precision. If all of the studies included in the analysis are estimating the same effect, the plot should display a distribution of effect sizes around the average effect, and the spread of effect sizes should be proportional to their variances. Therefore, small studies would be expected to be widely spread around the mean effect, with larger samples having a narrower spread. A statistical test assesses whether the degree of asymmetry (if any) in the funnel plot is more than would be expected by chance (Egger, Smith, Schneider, & Minder, 1997). More specifically, a Kendall's tau rank correlation test evaluates the relationship between the standardized effect sizes and their variances. If the test suggests a lack of symmetry around the average effect in the funnel plot, especially if small, negative studies are absent from the pool of included studies, publication bias is a genuine concern.

Publication bias can also be assessed through a regression asymmetry plot, in which the standardized effect size estimates are plotted against their inverse standard errors (Egger et al., 1997; Steichen, 1998). Egger's test examines whether the intercept in the plot deviates significantly from zero. According to Steichen (1998), as a result of their large standard errors small studies would be expected to have standardized effects and precision of close to zero, while large studies (due to their small standard errors) would have large standardized effects and high precision. When publication bias and asymmetry are evident, the regression line will deviate from zero because the set of small studies differs systematically from the set of larger studies. The further the intercept from zero, the greater the level of asymmetry in the plot.

**Solution sensitivity**

Given that the purpose of a meta-analysis is to summarize the findings from a body of studies assessing the same treatment effect, it is wise to consider the impact of outlier effect size values that differ substantially from the majority of the studies in the included set (Lipsey & Wilson, 2001). If present, extreme effect size values may exert a disproportionate influence on the pooled treatment effect across all included studies, which may impact the summative

---

76 Publication bias and small study effects are not distinguishable from one another by this test.

77 The Begg and Mazumdar test is reported to be underpowered, particularly for meta-analyses involving a small number of effect size estimates (Steichen, 1998).
meta-analytic conclusions and possibly lead to spurious results. An analysis of solution sensitivity involves examining the degree to which any individual study's effect size influences the pooled mean effect size across all studies. An appropriate approach is to inspect funnel plots for outlier studies that appear particularly influential with respect to effect size or sample size; in other words, examine the distribution of the effect sizes for the presence of outliers (Hedges & Olkin, 1985, Sweet & Applebaum, 2004). In addition, the pooled coefficient of the treatment effect can be tested for sensitivity through an iterative process that uses one-at-a-time removal of each of the individual studies from the analysis. If outliers are identified, their values can be adjusted to be more representative of the majority of the study estimates, or the outliers can simply be excluded from the analysis (Hedges & Olkin, 1985).

METHODS

Objectives of the Current Study

To examine the effects of school-based bullying prevention programs on the occurrence of bullying and victimization, this review identifies all relevant primary data evaluating such interventions and implements meta-analysis to summarize the findings from these studies. Specifically, this review is designed to assess if there is evidence for the following research question:

Are school-based bullying prevention programs associated with a reduction in the incidence of bullying and victimization?

The meta-analysis reported herein extends previous narrative reviews and meta-analytic efforts in several respects. First, unlike the bulk of the prior work the current study involves a quantitative assessment of treatment effectiveness as opposed to a narrative discussion and comparison of study outcomes. This is a significant extension of the literature on bullying prevention that provides a number of benefits such as greater precision in
estimating the treatment impact, greater objectivity, and so forth (see the section *Strengths of Meta-Analysis* for further discussion).

Compared with the meta-analysis by Ferguson and colleagues, the current study allows for the inclusion of studies that use rigorous quasi-experimental research designs as well as those that employ experimental research designs. This divergence from Ferguson substantially expands the pool of eligible studies while still maintaining a strict standard of methodological rigor for study inclusion. Furthermore, the literature search was not limited by publication date, and studies found in non-peer reviewed sources such as doctoral dissertations and books were eligible for inclusion in the analysis. Unlike Ferguson, the current study restricts outcomes to bullying and victimization outcomes only; findings related to more general aggressive or violent behavior are excluded from the analysis. This limitation allows for program impacts on bullying and victimization to be assessed in isolation from other treatment outcomes, and restricts outcomes to those that can truly be categorized as bullying behaviors. Finally, analyses herein are more comprehensive and involve the assessment of both bullying and victimization outcomes, as well as an exploration of between-study heterogeneity through stratification on study characteristics and treatment program elements.

Compared with the meta-analysis by Ttofi and Farrington (2009), the current study allows for samples with a minimum of 20 subjects per group (versus at least 200 subjects total), and includes studies that use self-reports or peer reports of bullying and victimization (versus restricting the set to self-reports only). Eight studies included in the current analysis were not used by Ttofi and Farrington; the discrepancy may be a result of this difference in study inclusion criteria. On the other hand, due to resource constraints the current analysis limited the set of included studies to those published in the English language; two studies were included by Ttofi and Farrington that were excluded in the current study as a result of this limitation.78 In addition, this study restricted methodological design to evaluations using either randomization to groups or a pretest/posttest independent-groups design, i.e., research

---

78 Due to being published in a non-English language, these two studies were not thoroughly reviewed and it is possible that upon closer inspection they would have been excluded for another reason.
designs such as age cohort control or quasi-experimental designs without a pretest comparison of group differences were not included. Ten studies included by Ttofi and Farrington were excluded from the current analysis due to not meeting this methodological requirement.79

Identification and Selection of Studies

Search Process

A systematic attempt was made to identify and retrieve all primary empirical studies meeting the eligibility criteria specified below. The primary source for such studies was a search of electronic bibliographic databases, from the inception of each database through April 23, 2008. These databases include:

- Criminal Justice Abstracts,
- Criminology: A SAGE Full-Text Collection,
- Dissertation Abstracts International,
- Education: A SAGE Full-Text Collection,
- Education Resources Information Center (ERIC),
- Medline,
- National Criminal Justice Reference Service Abstracts (NCJRS),
- Psychology: A SAGE Full-Text Collection,
- PsycInfo,
- SocINDEX,
- Sociological Abstracts,
- Sociology of Education Abstracts,
- Sociology: A SAGE Full-Text Collection, and

79 An additional three studies included by Ttofi and Farrington (2009) were excluded from the current meta-analysis due to not meeting other inclusion criteria.
The following keywords were used in an abstract search, and concepts were combined: ("peer victimization" or "peer harassment" or bully* or bullie*) AND (program or intervention or treatment or policy or policies) AND (evaluat* or assess*).

In addition, the bibliographies of previous meta-analyses and comprehensive literature reviews on bullying programs were reviewed (including Baldry & Farrington, 2007; Ferguson et al., 2007; Rigby, 2002a; Smith et al., 2003; Smith et al., 2004a; Ttofi & Farrington, 2009; and Vreeman & Carroll, 2007). The reference lists of the studies that met inclusion criteria were also reviewed for additional studies that might be candidates for inclusion in the current analysis. Over 95 percent of the studies identified as potentially eligible for inclusion in the analysis were retrieved and screened for eligibility.

Selection Criteria

An initial 2,330 titles/abstracts were identified by the primary search. Studies were retrieved for further review if the abstract or title suggested that the study might meet the inclusion criteria designated for the meta-analysis. Specifically, studies were included on the basis of the following six criteria:80,81

**Criteria relating to treatment program**

- The study evaluated a program, intervention, or policy in a primary, secondary, or K-12 school for students attending that school.

**Criteria relating to study outcomes**

- The study assessed at least one outcome variable that represented bullying behavior or bully victimization, including either perpetrating physical or non-physical bullying

---

80 Studies were not required to be published or peer-reviewed. While this decision may lack a degree of quality control, it also enables consideration of a large number of studies that would otherwise be automatically be excluded, and helps to counter the problem of publication bias. For example, many dissertations and theses remain unpublished, yet employ rigorous methodologies that are comparable to peer-reviewed journal articles (and have been subjected to peer review by dissertation committees). Furthermore, peer reviewers for journals are engaged to review written manuscripts without access to any other data than is accessible to the average reader. Since all studies included in this meta-analysis were carefully scrutinized with respect to methodological characteristics, this in itself can be viewed as an exercise in peer review/quality control.

81 Study inclusion was not limited by publication date.
on another student, or suffering physical or non-physical victimization by a bully. The definition of bullying or victimization used in the study must have been concordant with definitions used in bullying research (e.g., see Olweus, 1994, 1996; Rigby, 2004; Chapter 1 of the dissertation). Briefly, bullying involves an aggressive act (physical, verbal, or psychological) that is intended to cause harm or distress, and in which there is an imbalance of power or strength in the relationship between the more powerful bully/bullies and the less powerful victim/victims.82

- The outcome variable was measured using self-report questionnaires, peer nominations, teacher ratings, or direct observations of student behavior.
- The research report provided sufficient numerical or graphical data to allow for the calculation of an effect size with respect to treatment impact.83

Criteria relating to study protocol

- The number of research subjects in each experimental group was at least 20 students.
- The study employed one of two design types:
  (a) a randomized design in which subjects were assigned to an experimental or no-treatment control group, with or without inclusion of pretest measures, OR
  (b) an independent groups pretest-posttest design,84 with either random or non-random assignment to groups. In other words,
  - the study compared one or more experimental groups receiving one or more interventions with at least one comparison group not receiving any intervention, AND

---

82 The tri-criterion definition of bullying, as discussed in Chapter 1 of the dissertation, also requires that the aggressive behaviors occur repeatedly over time. This element of the definition was not required for inclusion in the current study, as many of the outcome assessments did not measure repetition of behaviors.

83 Several studies that otherwise met inclusion criteria did not permit this calculation (e.g., Smith, Sharp, Eslea & Thompson, 2001). For example, studies may not have reported sample sizes for treatment and control groups, or standard deviations of mean scores, without providing additional data to circumvent this requirement such as results from an ANOVA, t-test, chi-square, etc.

84 A repeated measures design on a single group was not considered sufficient, nor was an independent groups posttest design (see Cook & Campbell (1979) and Morris & DeShon (2002) for details on these designs and their limitations, for example, potential selection effects such as pre-existing group differences, time effects such as history or maturation bias, among others).
• the study conducted both pre- and post-intervention assessments of the same subjects on one or more bullying or victimization outcome measures.

In addition, research was excluded that:

• Examined the effect of a bullying intervention but did not allow for results to be isolated from the effects of another (known) ongoing intervention at the school, such as a general violence prevention program or a social skills improvement program.

• Examined the impact of a bullying intervention on student outcomes other than bullying or victimization (e.g., general aggression or violence) if bullying outcomes could not be assessed in isolation.

• Evaluated the impact of an intervention on attitudes (as opposed to behavior) regarding bullying, or anxiety or distress related to bullying involvement.

• Assessed program impacts on the bullying of or by individuals other than students, such as teachers or parents.

• Evaluated a bullying prevention program implemented outside of school hours or a school environment, such as an after-school program or a community center intervention.

• Employed a comparison group that was substantially different from the treatment group at baseline on pretest levels of bullying or victimization, and did not use rigorous statistical techniques to control for these differences. In other words, the comparison group and treatment group were deemed sufficiently different at baseline that they did not represent the same underlying population of students.
Independence of Effect Sizes

Following the identification and retrieval of the set of studies that meet inclusion criteria, another consideration is the independence of units of analysis. One source of nonindependence arises if the studies do not use independent populations, i.e., if multiple studies report on the same population of students. This was not an issue in the current analysis, as no two studies meeting inclusion criteria and assessing the same sample were identified through the systematic literature search.

Nonindependence is also a concern when a single study conducts more than one hypothesis test and presents results for multiple outcome assessments (Judd et al., 1991). Strategies for addressing this problem include pooling the results across the different outcome measures and then deriving a single effect size for the study, or ignoring the problem and including multiple outcomes per study and treating them as if they were independent (Judd et al., 1991). The current study uses another strategy: First, a clear differentiation was observed in that studies operationalized the dependent variable in terms of perpetration of bullying, victimization by bullying, or both. Therefore, the meta-analyses were split into bullying and victimization outcomes. Studies could provide effect sizes for one or both of the categories. Second, for studies that reported multiple operationalizations within the same category of dependent variable, for example, both physical and verbal bullying; only one outcome was used in the analysis. Several rules were applied to enable an objective decision with respect to the selection of effect sizes, as follows:

---

85 The exclusion of non-English studies was a potential source of bias for this meta-analysis. Egger et al. (1997) suggests that authors from non-English speaking countries may be more likely to report in an English-language journal if their results are statistically significant, and publish non-significant results in a local journal. If true, the likelihood of bias in the current meta-analysis is increased. On the other hand, Sterne et al. (1997) contend that the quality of studies published in languages other than English may be lower, which suggests that these studies typically would not meet the relatively strict requirements for inclusion in the current protocol. Despite these issues, due to lack of resources for translation the handful of non-English studies deemed potentially relevant based on abstracts were excluded from the current study.

86 A related issue that is a potential source for within-study bias is that multiple outcome measures may have been collected, but only one or a handful presented in the written report (Egger, Dickersin, & Smith, 2001). If this is the case, it is possible that authors will choose to report the most favorable findings with respect to their hypotheses (e.g., see Tannock, 1996). This type of bias is not observable in written reports and the extent to which it may have occurred is unknown.
**Decision rules for selecting among multiple effect sizes reported in a single study**

1. **Data source:** Where multiple sources of data were available (e.g., self-reports, peer nominations, or playground observations), self-report measures were used.\(^8^7\)

2. **Subgroups of outcome:** When results for a particular outcome type were separated into subgroups such as physical, relational or verbal bullying/victimization, physical bullying/victimization was selected.\(^8^8,^9^0\)

3. **Availability of data:** Where more than one outcome variable was reported (e.g., self-reports and peer nominations, or responses from primary and secondary school students), if one outcome was reported with greater precision it was selected for inclusion. This rule overrode Rules #1 and #2. For example, Turpeau (1998) presented self-reports of physical bullying/victimization, self-reports of verbal bullying/victimization, and peer reports of bullying/victimization. Although Rules #1 and #2 denote that self-reports of physical bullying/victimization should have been selected, peer nominations were chosen because the measure presented means and standard deviations. This data allowed for a more precise determination of the treatment effect than was possible using the self-report data that did not include standard deviations.\(^9^1\)

4. **Multiple posttests:** Where multiple posttest assessments were conducted, i.e., longer-term follow-ups, the first posttest was used in order to maintain greater homogeneity

---

\(^8^7\) For studies that reported bullying or victimization outcomes for separate age groups of students (i.e., younger versus older students), outcomes for both groups were included as separate effect sizes in the pooled analyses. More specifically, three studies contributed two effect sizes each, since the age-grouped subsamples contained within are independent. This is different from analyses of multiple outcomes on the same sample of students, for which using more than one outcome in the same meta-analysis would introduce statistical dependencies.

\(^8^8\) Affected 2 effect sizes for victimization outcomes: DeRosier (2004a) and Frey et al. (2005).

\(^8^9\) Affected 2 effect sizes for victimization outcomes: Bauer, Lozana, & Rivara (2007), and Woods, Coyle, Hoglund, & Leadbeater (2007).

\(^9^0\) Rules #1 and #2 jointly affected 1 effect size for bullying outcomes: Frey et al. (2005).

\(^9^1\) Affected 2 effect sizes for bullying outcomes and 3 effect sizes for victimization outcomes: Fekkes, Pijpers, & Verloove-Vanhorick (2006b; bullying and victimization); Jenson & Dieterich (2007; victimization); and Turpeau (1998; bullying and victimization).
among studies (because the majority of studies reported on a single posttest measure collected immediately after the conclusion of the intervention).\textsuperscript{92, 93}

The application of these decision rules did not introduce any obvious bias to the selection of treatment effects. For example, it was not the case that reports of physical bullying consistently resulted in stronger or weaker treatment effects than did reports of verbal bullying, or vice versa.

**Effect Size Calculations**

**Estimation of Missing Data**

For several of the studies included in the current analysis, pieces of information necessary to calculate the treatment effect size were not presented in the written report. In order to retain these studies in the analysis, some assumptions were made with respect to sample sizes and proportions, and the requisite data was extrapolated from the available study information. More specifically, one study (Melton et al., 1998) did not record the percentage of bullied and victimized students in numerical form, but provided graphical output that enabled simple and reasonably accurate estimates of the numerical data. Additionally, seven studies (Boulton & Flemington, 1996; Frey et al., 2005; Melton et al., 1998, Rosenbluth et al., 2004; Stevens et al., 2000; Taylor, 2006; and Woods et al., 2007) reported the total sample size of participants, but did not list the exact sample sizes in both treatment and control groups at pretest and posttest measurement points. For some of these studies, attrition rates or other related data on subject dropout between pre- and posttest were given for treatment or control groups, and therefore the number of subjects per group was fairly straightforward to estimate.

\textsuperscript{92} Affected 7 effect sizes for bullying outcomes: Cross, Hall, Hamilton, Pintabona, & Erecg (2004); DeRosier (2004a); Fekkes et al. (2006b); Menard et al. (2008); Rosenbluth, Whitaker, Sanchez, & Valle (2004); Stevens, De Bourdeaudhuij, & Van Oost (2000); and Taylor (2006).

\textsuperscript{93} Affect 8 effect sizes for victimization outcomes: Cross et al. (2004); DeRosier (2004a); Fekkes et al. (2006b); Menard et al. (2008); Rosenbluth et al. (2004); Stevens et al. (2000); Taylor (2006); and Woods et al. (2007).
using basic math. In the absence of any other data, the assumption was made that the group sizes were equivalent, or that attrition from pretest to posttest was equivalent for both groups.

**Effect Size Formulas**

Effect sizes in the current study were calculated using the following formulas, discussed in detail in the *Effect Size Calculations* section:

(a) For studies using continuous outcome measures and reporting standard deviations for the mean scores in the treatment and control groups, the standardized mean difference (SMD) effect size was derived using the following:

\[
ES_{SMD} = \left[1 - \frac{3}{4(n_T + n_C - 2)}\right] \frac{\left(\bar{Y}_{T,post} - \bar{Y}_{T,pre}\right) - \left(\bar{Y}_{C,post} - \bar{Y}_{C,pre}\right)}{SD_{post}},
\]

where \(n_T\) and \(n_C\) are posttest sample sizes for the treatment and control groups, respectively; \(\bar{Y}_{T,post}\) and \(\bar{Y}_{C,post}\) are the mean scores at posttest for the treatment and control groups, respectively; \(\bar{Y}_{T,pre}\) and \(\bar{Y}_{C,pre}\) are the mean scores at pretest for the treatment and control groups, and \(SD_{post}\) is the pooled post-test standard deviation for the treatment and control groups.

The pooled standard deviation is defined as follows:

\[
SD_{post} = \sqrt{\frac{(n_T - 1)SD_{T,post}^2 + (n_C - 1)SD_{C,post}^2}{n_T + n_C - 2}},
\]

where \(SD_{T,post}^2\) and \(SD_{C,post}^2\) are the squared standard deviation for the treatment group and control group at posttest, respectively.
The standard error of $ES_{SMD}$ is:

$$SE_{SMD} = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES_{SMD})^2}{2(n_T + n_C)}}.$$

Fifteen effect sizes for bullying outcomes and 16 effect sizes for victimization outcomes were estimated using this equation for standardized mean difference.

(b) One study in the current analysis, Menard et al. (2008), did not report mean scores or standard deviations for the treatment and control groups, but provided product moment correlations between the bullying outcome variable and the treatment/control group variable. For this study, the following conversion formula was used to estimate treatment effect size, resulting in one effect size for bullying outcomes and one effect size for victimization outcomes:

$$ES_r = \left[\frac{r_{post}}{\sqrt{(1-r_{post}^2)(p(1-p))}}\right] - \left[\frac{r_{pre}}{\sqrt{(1-r_{pre}^2)(p(1-p))}}\right],$$

where $r_{post}$ is the product moment correlation at posttest, $r_{pre}$ is the product moment correlation at pretest, and $p$ is the proportion of subjects in the treatment group.

The standard error is calculated using the following formula:

$$SE_r = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES)^2}{2(n_T + n_C)}}.$$
(c) For studies using dichotomous outcome measures and presenting raw numbers or percentages of treatment group and control group subjects who were bullied and/or victimized at pretest and posttest, the Cox-transformed log odds ratio was used to calculate effect size, as follows:

\[ ES_{Cox} = \left[ \frac{L_{OR,post}}{1.65} \right] - \left[ \frac{L_{OR,pre}}{1.65} \right] \]

where \( L_{OR,post} \) and \( L_{OR,pre} \) are the log odds ratios at posttest and pretest, respectively, \( p_T \) is the proportion in the treatment group that was not bullied (or victimized) and \( p_C \) is the proportion in the control group that was not bullied/victimized.

The standard error for the Cox-transformed odds ratio is:

\[ SE_{Cox} = \sqrt{0.367 \left[ \frac{1}{O_{1T}} + \frac{1}{O_{2T}} + \frac{1}{O_{1C}} + \frac{1}{O_{2C}} \right]} \]

where \( O_{1T} \) and \( O_{1C} \) are the successful outcome frequencies (i.e., non-bullied or non-victimized) in the treatment and control groups, respectively, and \( O_{2T} \) and \( O_{2C} \) are the non-successful outcome frequencies in the treatment and control groups.
Six of the effect sizes assessing bullying outcomes and eight of the effect sizes assessing victimization outcomes were calculated with the Cox-transformed formula for effect size.\textsuperscript{94}

Cluster Adjustments

The majority of the studies included in the current analysis did not involve assignment of individual students to treatment, but instead involved assignment of classrooms or schools to treatment or comparison groups. As such, a cluster adjustment of the effect sizes and standard errors was appropriate, since students were nested in classrooms and schools (see Hedges, 2007). This adjustment for clustering depends on an estimate of the intraclass correlation coefficient (ICC, or rho), a measure of the proportion of the variance attributable to group differences. As the intraclass correlation coefficient was not reported in any of the studies, a rho of 0.10 was used for the adjustment, as recommended by What Works Clearinghouse (2008) for studies of behavioral outcomes.

Other Data Coded

Along with the data used to calculate effect sizes, some additional information from each study was coded. This included details with respect to study design (e.g., how subjects were selected and assigned to experimental groups), geographic location (North America, Europe, or Australia), type of publication (e.g., journal article, book chapter), year of publication, participant age, and specifics on program content (e.g., curriculum-based, whole-school anti-bullying policy, information for parents, and so forth). The characteristics of the bullying prevention programs are shown in Table 30.

\textsuperscript{94} For one study, Jenson and Dieterich (2007), the Cox-transformed log odds ratio was computed but it was not possible to compute the Cox standard error because the authors did not present 2x2 cell frequencies for successful/non-successful treatment outcomes. For this study, the standard error was computed using the formula for $ES_{Hedges}$, which is not perfect but is the next best estimate (as recommended by M. Lipsey, personal communication, February 22, 2009).
Software

Meta-analyses and meta-regressions were conducted using Stata SE 9.2 (Stata Corporation, College Station: Texas). Adjustments of effect sizes and standard errors for clustering were achieved using a spreadsheet calculator developed by McHugh (2004).

RESULTS

After applying the inclusion and exclusion criteria as outlined in the meta-analysis protocol, 23 studies met the requirements and were included in the analysis. Analyses of the effects of prevention programs were undertaken separately for the outcomes of bullying (i.e., the perpetration of bullying) and victimization (being bullied by others). A total of 47 effect sizes were retrieved, with 22 describing bullying outcomes and 25 describing victimization outcomes.

Characteristics of the Included Studies

Twenty-two independent effect sizes assessing the effects of a bullying prevention program on bullying outcomes were extracted from 19 studies involving 18,903 students. For victimization outcomes, 25 independent effect sizes were extracted from 22 studies involving 25,361 students. Table 29 presents descriptive data on the characteristics of the studies included in the analysis. The majority of the studies reported both bullying and victimization outcomes and were included in both analyses; only 5 studies (1 bullying and 4 victimization) did not report on both types of outcomes. Despite not restricting the time frame for study inclusion, the majority of the studies were conducted within the past decade.

---

95 The data for the bullying and victimization effect sizes are presented as a means for describing the types of studies included in each of the meta-analyses and are not intended to be compared with one another.

96 Results for bullying and victimization outcomes are presented as summaries of the independent effect sizes used in each analysis, not as summaries of the individual studies. More specifically, three of the studies contributed two independent effect sizes to each analysis (using samples of older and younger students); as such they are each counted twice in these descriptive outcomes (Baldry & Farrington, 2004; Rahey & Craig, 2002; and Stevens et al., 2000).

97 Bullying only: Boulton & Flemington, 1996. Victimization only: Alsaker & Valkanover, 2001; Bauer et al., 2007; Beran, Tutty, & Steinrath, 2004; and Woods et al., 2007.
The date range for publication was 1996 to 2008, with 82 percent of the bullying studies and 88 percent of the victimization studies published since 2000. The majority of the studies were published in peer-reviewed sources, including academic journals (68 percent for bullying and 64 percent for victimization), or chapters in an edited book (9 percent for bullying and 16 percent for victimization). Studies presented in non-peer-reviewed sources include doctoral dissertations (14 percent for bullying and 12 percent for victimization), and technical reports (9 percent bullying and 8 percent victimization). Fifty percent of the bullying studies and 56 percent of the victimization studies were conducted in the United States or Canada. The remaining studies were mostly set in Europe (41 percent of the bullying and 36 percent of the victimization studies), with 9 percent of the bullying and 8 percent of the victimization studies taking place in Australia.

Close to half of the studies used some form of randomized assignment to treatment and control groups (55 percent of bullying and 44 percent of victimization studies), while 45 percent of the bullying and 56 percent of the victimization studies used a quasi-experimental design with nonrandom assignment. The most common unit of assignment to treatment condition was schools (46 percent bullying and 52 percent victimization); followed by assignment of classrooms to treatment condition (41 percent bullying and 36 percent victimization). Ninety-one percent of the bullying studies and 88 percent of the victimization studies used self-reports of bullying or victimization outcomes, while the remaining effect sizes were estimated from outcome measures using peer reports. Sample sizes varied dramatically among studies with a range from 43 to 6,389 students in both treatment and control groups at pretest. Sample size was dichotomized for the purposes of analysis, with 55 percent of the bullying studies and 52 percent of the victimization studies using a total sample of less than 400 students, and 46 percent of the bullying and 48 percent of the victimization studies using samples of 400 students or more.
Table 29 Characteristics of Studies Included in the Meta-Analyses

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Bullying outcomes (n=22)</th>
<th>Victimization outcomes (n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996-1999</td>
<td>4 (18.2%)</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td>2000-2003</td>
<td>5 (22.7%)</td>
<td>6 (24.0%)</td>
</tr>
<tr>
<td>2004-2008</td>
<td>13 (59.1%)</td>
<td>16 (64.0%)</td>
</tr>
<tr>
<td>Publication type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer-reviewed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal article</td>
<td>15 (68.2%)</td>
<td>16 (64.0%)</td>
</tr>
<tr>
<td>Book chapter</td>
<td>2 (9.1%)</td>
<td>4 (16.0%)</td>
</tr>
<tr>
<td>Non-peer-reviewed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissertation</td>
<td>3 (13.6%)</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td>Report</td>
<td>2 (9.1%)</td>
<td>2 (8.0%)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States or Canada</td>
<td>11 (50.0%)</td>
<td>14 (56.0%)</td>
</tr>
<tr>
<td>Europe</td>
<td>9 (40.9%)</td>
<td>9 (36.0%)</td>
</tr>
<tr>
<td>Australia</td>
<td>2 (9.1%)</td>
<td>2 (8.0%)</td>
</tr>
<tr>
<td>Type of research design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random after matching/stratification</td>
<td>5 (22.7%)</td>
<td>5 (20.0%)</td>
</tr>
<tr>
<td>Random simple</td>
<td>7 (31.8%)</td>
<td>6 (24.0%)</td>
</tr>
<tr>
<td>Non-randomized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonrandom, post hoc matching</td>
<td>3 (13.6%)</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td>Nonrandom, self-selection</td>
<td>7 (31.8%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>Unit of assignment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>10 (45.5%)</td>
<td>13 (52.0%)</td>
</tr>
<tr>
<td>Classroom</td>
<td>9 (40.9%)</td>
<td>9 (36.0%)</td>
</tr>
<tr>
<td>District</td>
<td>1 (4.6%)</td>
<td>1 (4.0%)</td>
</tr>
<tr>
<td>Student</td>
<td>2 (9.1%)</td>
<td>2 (8.0%)</td>
</tr>
<tr>
<td>Outcome source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report</td>
<td>20 (90.9%)</td>
<td>22 (88.0%)</td>
</tr>
<tr>
<td>Peer report</td>
<td>2 (9.1%)</td>
<td>3 (12.0%)</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 400 students</td>
<td>12 (54.6%)</td>
<td>13 (52.0%)</td>
</tr>
<tr>
<td>400+ students</td>
<td>10 (45.5%)</td>
<td>12 (48.0%)</td>
</tr>
</tbody>
</table>

Table 30 presents basic information about the characteristics of the bullying prevention programs. Around half of the programs were implemented at a comprehensive, whole school level, while 45 percent of the programs were implemented at the individual classroom level.
One study was a small group intervention for victims of bullying (4 percent). The duration of the bullying prevention programs ranged from a single session to two academic years, with an average program length across all 22 bullying studies of 22.9 weeks, and an average program length of 27.6 weeks across the 25 victimization studies. The age range of participating students was 5 to 18 years, with an average of 11.4 years for the bullying studies and 10.9 years for the victimization studies. Eight-two percent of the programs assessing bullying outcomes and 84 percent of the victimization programs used a specific classroom curriculum as part of the bullying prevention strategy, while the remaining programs used unstructured anti-bullying strategies in the classroom or did not use any classroom strategies at all. Anti-bullying rules were commonly used, with 41 percent of bullying and 44 percent of victimization programs implementing a formal anti-bullying policy at a whole school level, and 46 percent of bullying and 52 percent of victimization programs implementing anti-bullying rules at the classroom level. Most of the programs (68 percent of bullying and 76 percent of victimization) involved teacher training with respect to bullying prevention and/or curriculum delivery. Fifty percent of the bullying programs and 44 percent of the victimization programs incorporated individualized work with students identified as bullies or victims (often provided by trained professionals, such as psychologists). Around a third of the programs involved an element of peer engagement through either peer mentoring or peer mediation. Finally, 41 percent of the bullying programs and 52 percent of the victimization programs involved parents to some capacity, through informational newsletters or pamphlets regarding the anti-bullying initiative, at-home manuals to assist them in helping their child deal with bullying situations at school, or through face-to-face parent/teacher meetings or anti-bullying presentations at the school.

---

98 For example, a one-time video intervention followed by a short discussion without any additional anti-bullying activities or emphasis during classroom lessons.

99 Often, classroom rules were developed by the class itself as part of the treatment program, with students and teachers working together to decide on appropriate rules to prevent bullying behaviors.

100 For example, students selected to work as mediators in the aftermath of a bullying incident, or older students acting as peer mentors to younger students involved in bullying or victimization.
Table 30 Characteristics of the School-Based Bullying Prevention Programs

<table>
<thead>
<tr>
<th>Program characteristics</th>
<th>Bullying outcomes (n=22)</th>
<th>Victimization outcomes (n=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of implementation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole school</td>
<td>11 (50.0%)</td>
<td>13 (52.0%)</td>
</tr>
<tr>
<td>Classroom</td>
<td>10 (45.5%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>Small group (targeted intervention)</td>
<td>1 (4.5%)</td>
<td>1 (4.0%)</td>
</tr>
<tr>
<td>Mean no. of weeks in length</td>
<td>22.94 weeks (range 0.2-72)</td>
<td>27.62 weeks (range 0.2-72)</td>
</tr>
<tr>
<td>Program duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 weeks or less</td>
<td>11 (50.0%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>More than 12 weeks</td>
<td>11 (50.0%)</td>
<td>14 (56.0%)</td>
</tr>
<tr>
<td>Mean participant age</td>
<td>11.39 years (range 6-18)</td>
<td>10.92 years (range 5-18)</td>
</tr>
<tr>
<td>Mean participant age (dichotomized)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 11 years</td>
<td>11 (50.0%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>11 years or older</td>
<td>11 (50.0%)</td>
<td>14 (56.0%)</td>
</tr>
<tr>
<td>Included program elements (dichotomous indicators of yes/no)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum</td>
<td>18 (81.8%)</td>
<td>21 (84.0%)</td>
</tr>
<tr>
<td>Whole-school anti-bullying rules</td>
<td>9 (40.9%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>Classroom anti-bullying rules</td>
<td>10 (45.5%)</td>
<td>13 (52.0%)</td>
</tr>
<tr>
<td>Teacher training</td>
<td>15 (68.2%)</td>
<td>19 (76.0%)</td>
</tr>
<tr>
<td>Individual work with bullies and/or victims</td>
<td>11 (50.0%)</td>
<td>11 (44.0%)</td>
</tr>
<tr>
<td>Peer mediation or mentoring</td>
<td>8 (36.4%)</td>
<td>8 (32.0%)</td>
</tr>
<tr>
<td>Information or meetings for parents</td>
<td>9 (40.9%)</td>
<td>13 (52.0%)</td>
</tr>
</tbody>
</table>

Table 44 in Appendix 2 outlines key characteristics of the 23 studies included in the analyses, while Table 45 in Appendix 2 lists treatment program elements for each of the 23 studies.

Meta-Analysis of Bullying Outcomes

Pooled Results

The 22 effect sizes assessing bullying outcomes are reasonably homogenous, with between-study variance as would be expected from sampling error. Calculations for the $Q$-statistic are presented in Appendix 4; these show that the $Q$-statistic for homogeneity of 11.64
Figure 4 presents a Galbraith plot, which displays the extent of heterogeneity among effect sizes included in the meta-analysis (Galbraith, 1988; Tobias, 1998). For each effect size, represented by a dot on the plot, the Z-statistic ($Z = \frac{ES}{SE_{ES}}$) is plotted against the inverse standard error. The center line in the plot represents the unweighted regression line (constrained through the origin), with the parallel lines representing two standard deviations above and below the regression estimate. These two lines represent the 95% confidence interval within which most of the study estimates would be expected to fall if the studies were similar enough to be estimating a fixed parameter (Bax et al., 2009). If effect sizes are substantially heterogeneous, some data points would be expected to fall outside the confidence bounds. As indicated by both the Galbraith plot and the $Q$-statistic, negligible heterogeneity was present among the 22 effect sizes included in the bullying meta-analysis; as such, the fixed effects model was used to estimate the pooled treatment effect. To illustrate the derivation of the pooled effect size in the fixed effects model, calculations are provided in Appendix 3.

---

101 $Q$ is distributed as a chi-square with degrees of freedom equal to $k - 1$, where $k$ is the number of effect sizes.
102 The $Q$-statistic of 11.64 also does not exceed $Q = 29.6$ at $p < .10$, $df = 21$. Some authors, such as Higgins et al. (2003), suggest using a 10% significance rate to help address the concern of $Q$ being underpowered.
Table 31 presents the 22 effect sizes included in the meta-analysis on bullying outcomes. According to the conventions put forth by Cohen (1988), nearly all of the effect sizes are small (n=21), with one effect size of .553 in the medium range. Table 31 also presents estimates of the corresponding standard error, variance, inverse variance weight, and relative weight\textsuperscript{103} for each effect size using the fixed effects model.

---

\textsuperscript{103} Relative weight is the percentage of the sum of weights attributed to each individual study.
Table 31 Study-Level Data for the Meta-Analysis of Bullying Outcomes

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Effect size (ES)</th>
<th>Standard error (SE)</th>
<th>Variance (v)</th>
<th>Inverse variance weight (w)</th>
<th>Relative weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Andreou (2007)</td>
<td>0.355</td>
<td>0.197</td>
<td>0.039</td>
<td>25.64</td>
<td>7.8%</td>
</tr>
<tr>
<td>2 Baldry #1 (2004)</td>
<td>-0.376</td>
<td>0.297</td>
<td>0.088</td>
<td>11.36</td>
<td>3.5%</td>
</tr>
<tr>
<td>3 Baldry #2 (2004)</td>
<td>0.446</td>
<td>0.274</td>
<td>0.075</td>
<td>13.33</td>
<td>4.1%</td>
</tr>
<tr>
<td>4 Boulton (1996)</td>
<td>0.075</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>4.3%</td>
</tr>
<tr>
<td>5 Cross (2004)</td>
<td>-0.289</td>
<td>0.421</td>
<td>0.177</td>
<td>5.64</td>
<td>1.7%</td>
</tr>
<tr>
<td>6 DeRosier (2004a)</td>
<td>0.102</td>
<td>0.166</td>
<td>0.028</td>
<td>36.23</td>
<td>11.0%</td>
</tr>
<tr>
<td>7 Fekkes (2006b)</td>
<td>0.073</td>
<td>0.271</td>
<td>0.073</td>
<td>13.62</td>
<td>4.1%</td>
</tr>
<tr>
<td>8 Frey (2005)</td>
<td>0.059</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>4.3%</td>
</tr>
<tr>
<td>9 Hunt (2007)</td>
<td>-0.110</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>3.1%</td>
</tr>
<tr>
<td>10 Jenson (2007)</td>
<td>0.098</td>
<td>0.142</td>
<td>0.020</td>
<td>49.75</td>
<td>15.1%</td>
</tr>
<tr>
<td>11 Melton (1998)</td>
<td>0.217</td>
<td>0.390</td>
<td>0.152</td>
<td>6.57</td>
<td>2.0%</td>
</tr>
<tr>
<td>12 Menard (2008)</td>
<td>0.083</td>
<td>0.272</td>
<td>0.074</td>
<td>13.51</td>
<td>4.1%</td>
</tr>
<tr>
<td>13 Menesini (2003)</td>
<td>0.255</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>6.8%</td>
</tr>
<tr>
<td>14 Mikheev (2005)</td>
<td>0.459</td>
<td>0.310</td>
<td>0.096</td>
<td>10.42</td>
<td>3.2%</td>
</tr>
<tr>
<td>15 Rahey #1 (2002)</td>
<td>0.097</td>
<td>0.470</td>
<td>0.221</td>
<td>4.52</td>
<td>1.4%</td>
</tr>
<tr>
<td>16 Rahey #2 (2002)</td>
<td>0.036</td>
<td>0.460</td>
<td>0.212</td>
<td>4.72</td>
<td>1.4%</td>
</tr>
<tr>
<td>17 Rican (1996)</td>
<td>0.553</td>
<td>0.559</td>
<td>0.312</td>
<td>3.20</td>
<td>1.0%</td>
</tr>
<tr>
<td>18 Rosenbluth (2004)</td>
<td>-0.006</td>
<td>0.303</td>
<td>0.092</td>
<td>10.89</td>
<td>3.3%</td>
</tr>
<tr>
<td>19 Stevens #1 (2000)</td>
<td>0.174</td>
<td>0.232</td>
<td>0.054</td>
<td>18.52</td>
<td>5.6%</td>
</tr>
<tr>
<td>20 Stevens #2 (2000)</td>
<td>-0.169</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>6.8%</td>
</tr>
<tr>
<td>21 Taylor (2006)</td>
<td>0.088</td>
<td>0.359</td>
<td>0.129</td>
<td>7.75</td>
<td>2.4%</td>
</tr>
<tr>
<td>22 Turpeau (1998)</td>
<td>0.096</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>3.1%</td>
</tr>
<tr>
<td>Σ</td>
<td>2.316</td>
<td>6.707</td>
<td>2.271</td>
<td>328.71</td>
<td>100%</td>
</tr>
</tbody>
</table>

The pooled meta-analysis estimate, 95% confidence interval, and Z-score are presented in Table 32. The pooled estimate of the treatment effect on bullying outcomes is 0.109 (95% CI = 0.001, 0.217, Z=1.98, p=.047). Since the Z-score of 1.98 exceeds the critical value of 1.96 at p=.05, we conclude that the effect of bullying prevention programs on bullying outcomes is statistically significant.

---

104 Portions of the data for two of the included studies were obtained from the meta-analysis by Ttofi, Farrington, and Baldry (2008), as these data were unavailable in the published reports themselves. Specifically, standard deviations were obtained for the study by Baldry and Farrington (2004), and means and standard deviations were obtained for the study by Rahey and Craig (2002).
Table 32 Fixed Effects Meta-Analysis for Bullying Outcomes

<table>
<thead>
<tr>
<th>Pooled estimate</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>Z</th>
<th>p-value</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.109</td>
<td>0.001</td>
<td>0.217</td>
<td>1.983</td>
<td>0.047</td>
<td>22</td>
</tr>
</tbody>
</table>

Test for heterogeneity: $Q = 11.64$ on 21 degrees of freedom ($p = 0.949$), $I^2 = 0\%$

Figure 5 presents a forest plot of the study-level effect sizes for bullying outcomes. In the forest plot, the squares and the horizontal lines correspond to the studies' effect size estimates and 95% confidence intervals, respectively. The relative weight that each study contributes to the pooled analysis is represented by the area (size) of the square, while the length of the horizontal line represents the width of the confidence interval. The diamond at the bottom of the figure is the pooled estimate of the effect size (0.109) with its 95% confidence intervals indicated by the left and right edges of the diamond. As the left edge of the diamond is just over the zero point, the pooled estimate is deemed statistically significant.

Figure 5 Forest Plot of Study Effect Sizes for Bullying Outcomes

a Studies sorted by standard error
Investigating Between-Study Heterogeneity

Although the $Q$-statistic for the set of bullying effect sizes was not statistically significant, the possibility of systematic between-study heterogeneity was explored via moderator analysis. Some authors claim that $Q$ is underpowered, and that a non-significant $Q$-statistic does not necessarily preclude the investigation of potential moderators of treatment effects (e.g., Higgins et al., 2003; Lipsey & Wilson, 2001). Potential sources of the excess variability among studies were explored through subgroup analyses based on observable study and treatment characteristics, an approach that assumes a portion of the heterogeneity between studies is associated with particular variables that can be modeled in the analysis. The remaining variability is assumed to be random and therefore not suitable for modeling.

The approach used to assess variability in between-studies effects is an analog to the ANOVA (for more details, see the section Subgroup analyses). With this technique, the total variation is partitioned into the segment that is explained by between-group variation in the study/treatment characteristic of interest ($Q_B$), and the residual portion that remains ($Q_W$, an index of the variability within the group) (Lipsey & Wilson, 2001). If $Q_B$ is statistically significant, the categorical variable is assumed to systematically explain at least a portion of the between-groups variability. If, in addition, $Q_W$ is not statistically significant, the categorical variable is thought to account for all of the excess variability across the set of effect sizes – in other words, subject-level sampling error explains any remaining variation among effect sizes (Lipsey & Wilson, 2001).

To investigate potential moderator effects, fifteen dichotomous indicator variables were created: five variables addressing general study characteristics, and ten indicators reflecting elements of the treatment programs. The general study characteristic variables such as research design type, geographic location, and sample size were selected based on previous meta-analytic research that suggests these factors may be related to study outcomes (e.g., see Farrington & Welsh, 2003; Lipsey & Wilson, 2001; Sweet & Applebaum, 2004; Wilson & Lipsey, 2000). The treatment program variables were selected following a review of the general school bullying and bullying prevention literature; in which these factors were hypothesized to
be related to successful treatment outcomes (e.g., see Croft, 1996; Eslea & Smith, 1998; Espelage, 2004; Hazler & Carney, 2006; Olweus, 1994; Smith et al., 2004a; Vreeman & Carroll, 2007). The 15 variables include the following:

- **Study characteristics:**  
  1. Publication type (peer reviewed versus non-peer-reviewed)
  2. Geographic location (North America versus Europe/Australia)
  3. Research design (randomized versus non-randomized)
  4. Scale of outcome measure (continuous versus dichotomous)  
  5. Sample size (<400 students versus 400+ students)

- **Treatment elements:**
  1. Level of implementation (whole school versus classroom)
  2. Program duration (≤12 weeks versus >12 weeks)
  3. Mean participant age (<11 years versus 11+ years)
  4. Explicit classroom curriculum (yes/no)
  5. Whole school anti-bullying rules/policy (yes/no)
  6. Classroom-level anti-bullying rules (yes/no)
  7. Teacher training (yes/no)
  8. Individual work with bullies and/or victims (yes/no)
  9. Peer mediation/mentoring (yes/no)
  10. Information/meetings for parents (yes/no)

Table 33 presents results from the subgroup analyses of study characteristics on bullying outcomes. Subgroup estimates were compared using the analog to the ANOVA method to assess whether the categorical variable groupings help explain any of the excess

---

105 Outcome source (peer report versus self-report) was also planned for analysis; however, not enough variation was observed in the set of included effect sizes, i.e., the vast majority of the studies used self-reports of bullying and/or victimization.

106 Lipsey and Wilson (2001) suggest that if outcome variables assessed using continuous and dichotomous measures are combined into a pooled analysis, the scale of the outcome measure should be used in a moderator analysis to test whether this artifact of research design has an impact on effect size.

107 The study by DeRosier (2004a), in which the program was implemented outside of the classroom with small groups of victims, was dropped from this subgroup analysis.
variability in effect sizes. Measures of $Q_w$ (within-group heterogeneity) and $Q_b$ (between-groups heterogeneity) were calculated. See Appendix 5 for a demonstration of the analysis procedure using the variable "research design".

No significant measures of between-groups heterogeneity were found for any of the five study characteristic variables, suggesting that the mean effect sizes produced by each of these subgroups are not significantly different from one another. In other words, study characteristics in the form of publication type, geographic location, research design, scale of outcome measure, and sample size were not significant factors in predicting the magnitude of treatment effect on bullying outcomes.
Table 33 Subgroup Analysis of Study Characteristics on Bullying Outcomes

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Effect size and $Q$-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publication type</strong></td>
<td></td>
</tr>
<tr>
<td>Peer-reviewed (n=17)</td>
<td>$ES = 0.096$</td>
</tr>
<tr>
<td>Non peer-reviewed (n=5)</td>
<td>$ES = 0.185$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.33 \sim \chi^2_{11}, p=.566$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.31 \sim \chi^2_{20}, p=.938$</td>
</tr>
<tr>
<td><strong>Geographic location</strong></td>
<td></td>
</tr>
<tr>
<td>North America (n=11)</td>
<td>$ES = 0.112$</td>
</tr>
<tr>
<td>Europe or Australia (n=11)</td>
<td>$ES = 0.106$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.00 \sim \chi^2_{11}, p=.956$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.64 \sim \chi^2_{20}, p=.928$</td>
</tr>
<tr>
<td><strong>Research design</strong></td>
<td></td>
</tr>
<tr>
<td>Randomized (n=12)</td>
<td>$ES = 0.053$</td>
</tr>
<tr>
<td>Non-randomized (n=10)</td>
<td>$ES = 0.219$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 2.02 \sim \chi^2_{11}, p=.155$</td>
</tr>
<tr>
<td>Within-study heterogeneity</td>
<td>$Q_W = 9.62 \sim \chi^2_{20}, p=.975$</td>
</tr>
<tr>
<td><strong>Scale of outcome measure</strong></td>
<td></td>
</tr>
<tr>
<td>Dichotomous (n=6)</td>
<td>$ES = 0.082$</td>
</tr>
<tr>
<td>Continuous (n=16)</td>
<td>$ES = 0.120$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.09 \sim \chi^2_{11}, p=.762$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.55 \sim \chi^2_{20}, p=.931$</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;400 students (n=12)</td>
<td>$ES = 0.154$</td>
</tr>
<tr>
<td>400+ students (n=10)</td>
<td>$ES = 0.069$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.60 \sim \chi^2_{11}, p=.438$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.04 \sim \chi^2_{20}, p=.945$</td>
</tr>
</tbody>
</table>

Table 34 presents results from the subgroup analyses of ten treatment program elements on bullying outcomes. Again, no significant measures of between-groups heterogeneity were found, suggesting that none of the ten treatment elements are significant factors in predicting the magnitude of treatment effect on bullying outcomes.
<table>
<thead>
<tr>
<th>Treatment element</th>
<th>Effect size and $Q$-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Level of implementation</td>
<td></td>
</tr>
<tr>
<td>Whole school (n=11)</td>
<td>$ES = 0.015$</td>
</tr>
<tr>
<td>Classroom (n=10)</td>
<td>$ES = 0.177$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 1.88 \sim \chi^2_1$, $p=.171$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 9.76 \sim \chi^2_{19}$, $p=.959$</td>
</tr>
<tr>
<td>2 Program duration</td>
<td></td>
</tr>
<tr>
<td>$\leq 12$ weeks (n=11)</td>
<td>$ES = 0.123$</td>
</tr>
<tr>
<td>$&gt;12$ weeks (n=11)</td>
<td>$ES = 0.099$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.05 \sim \chi^2_1$, $p=.828$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.59 \sim \chi^2_{20}$, $p=.929$</td>
</tr>
<tr>
<td>3 Mean participant age</td>
<td></td>
</tr>
<tr>
<td>$&lt;11$ years (n=11)</td>
<td>$ES = 0.092$</td>
</tr>
<tr>
<td>$11+$ years (n=11)</td>
<td>$ES = 0.123$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.08 \sim \chi^2_1$, $p=.780$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.56 \sim \chi^2_{20}$, $p=.930$</td>
</tr>
<tr>
<td>4 Curriculum</td>
<td></td>
</tr>
<tr>
<td>Used curriculum (n=18)</td>
<td>$ES = 0.092$</td>
</tr>
<tr>
<td>No curriculum (n=4)</td>
<td>$ES = 0.215$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.60 \sim \chi^2_1$, $p=.438$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.04 \sim \chi^2_{20}$, $p=.945$</td>
</tr>
<tr>
<td>5 Whole-school anti-bullying policy</td>
<td></td>
</tr>
<tr>
<td>Whole-school policy (n=9)</td>
<td>$ES = 0.009$</td>
</tr>
<tr>
<td>No whole-school policy (n=13)</td>
<td>$ES = 0.164$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 1.79 \sim \chi^2_1$, $p=.180$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 9.84 \sim \chi^2_{20}$, $p=.971$</td>
</tr>
<tr>
<td>6 Classroom anti-bullying rules</td>
<td></td>
</tr>
<tr>
<td>Classroom rules (n=10)</td>
<td>$ES = 0.129$</td>
</tr>
<tr>
<td>No classroom rules (n=12)</td>
<td>$ES = 0.087$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.14 \sim \chi^2_1$, $p=.706$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.50 \sim \chi^2_{20}$, $p=.932$</td>
</tr>
<tr>
<td>7 Teacher training</td>
<td></td>
</tr>
<tr>
<td>Teacher training (n=15)</td>
<td>$ES = 0.095$</td>
</tr>
<tr>
<td>No teacher training (n=7)</td>
<td>$ES = 0.129$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.09 \sim \chi^2_1$, $p=.760$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.55 \sim \chi^2_{20}$, $p=.931$</td>
</tr>
<tr>
<td>8 Individual work with bullies and/or victims</td>
<td></td>
</tr>
<tr>
<td>Individual bully/victim work (n=11)</td>
<td>$ES = 0.070$</td>
</tr>
<tr>
<td>No individual bully/victim work (n=11)</td>
<td>$ES = 0.146$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.48 \sim \chi^2_1$, $p=.488$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 11.16 \sim \chi^2_{20}$, $p=.942$</td>
</tr>
</tbody>
</table>
Peer mediation or mentoring

<table>
<thead>
<tr>
<th>Condition</th>
<th>ES</th>
<th>QB</th>
<th>QB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer mediation or mentoring (n=8)</td>
<td>0.057</td>
<td>0.36 (\chi^2_1), p=.547</td>
<td></td>
</tr>
<tr>
<td>No peer mediation or mentoring (n=14)</td>
<td>0.131</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Within-studies heterogeneity

<table>
<thead>
<tr>
<th>Condition</th>
<th>ES</th>
<th>QB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-studies heterogeneity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parent information or meetings

<table>
<thead>
<tr>
<th>Condition</th>
<th>ES</th>
<th>QB</th>
<th>QB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents involved (n=9)</td>
<td>0.009</td>
<td>1.79 (\chi^2_1), p=.180</td>
<td></td>
</tr>
<tr>
<td>No parents involved (n=13)</td>
<td>0.164</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Within-studies heterogeneity

<table>
<thead>
<tr>
<th>Condition</th>
<th>ES</th>
<th>QB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-studies heterogeneity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitivity Analysis

Publication bias

As mentioned in the section *Weaknesses of Meta-Analysis*, publication bias is the elevated publication frequency of manuscripts containing statistically significant results compared with reports of non-significant or null results. If publication bias is present, such bias may distort the pooled effect size estimate in a meta-analysis (because reports with non-significant outcomes will not be included in the data set). Small effect sizes are often produced by studies with non-statistically significant findings or studies that were completed but remain unpublished (Lipsey & Wilson, 2001).

The analysis of publication bias includes the Begg and Mazumdar (1994) adjusted rank correlation test along with the Egger regression asymmetry test (Egger et al., 1997). The Begg and Mazumdar test uses a Kendall's tau adjusted rank correlation (Steichen, 1998) to examine whether there is a significant correlation between the (standardized) effect size estimates and their variances. The Begg and Mazumdar test plots each study's effect size against its standard error. In the absence of publication bias, the plot should resemble an inverted funnel with the vertex over the central effect estimate (Begg, 1994). Conversely, graphs with a positive skew suggest publication bias, as small effect sizes may be missing from such data sets.
The Egger test examines the level of funnel plot asymmetry by testing whether the regression line intercept deviates significantly from zero (Steichen, 1998). Table 35 and Figures 6 and 7 present the assessment of publication bias for the meta-analysis of bullying outcomes.

Table 35 Tests for Publication Bias in Bullying Outcomes Data Set

<table>
<thead>
<tr>
<th>Begg &amp; Mazumdar's Test</th>
<th>Adj. Kendall's Score (adj. for ties)</th>
<th>Std. dev. (adj. for ties)</th>
<th>No. of studies</th>
<th>Z (p-value) (continuity corrected)</th>
<th>Z (p-value) (continuity corrected)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-8</td>
<td>35.42</td>
<td>22</td>
<td>-0.23 (0.821)</td>
<td>0.20 (0.843)</td>
<td></td>
</tr>
</tbody>
</table>

| Egger's Test | Std effect | Coefficient | Std. error | t (p>|t|) | 95% Conf. Interval Lower | Upper |
|--------------|------------|-------------|------------|--------|-------------------------|-------|
| Slope        | .141       | .130        | 1.08       | .291   | -.130                   | .413  |
| Bias         | -.130      | .503        | -0.26      | .799   | -1.18                   | .920  |

Figure 6 Begg's Funnel Plot of Study Effect Size vs. Standard Error, Bullying Outcomes
The horizontal line across the funnel plot shows the inverse-variance-weighted fixed-effects treatment estimate. The two sloping lines represent the pseudo 95% confidence intervals for a given standard error, in other words, the expected confidence intervals if there was no heterogeneity among the studies. There is no evidence of publication bias using Begg's test ($Z=-0.23, p=0.821$), and the funnel plot appears reasonably symmetric.

**Figure 7 Egger's Publication Bias Plot of Standardized Effect Size Estimates vs. Precision, Bullying Outcomes**

The upper line in the Egger plot depicts the regression of standardized effect estimates on precision. There is no evidence of publication bias using the Egger test ($t$-statistic for bias$= -0.26, p=0.799$), indicating that the regression line intercept does not deviate significantly from zero.
Solution sensitivity

Solution sensitivity is concerned with whether any individual effect size appears to have a disproportionate influence on the pooled mean effect size. This was approached in two ways. First, the bullying outcomes funnel plot (Figure 6) was inspected for whether any individual data points (study effect sizes) fell outside of the pseudo-95% confidence intervals. No data points were identified that appeared abnormal.

Second, an influence analysis was undertaken in which the meta-analysis estimates were computed omitting each study in turn, which demonstrates the influence of each study on the overall pooled estimate (Tobias, 1999). In other words, this analysis demonstrates whether any individual effect size or effect sizes may be exerting a disproportionate influence on the pooled effect estimate. Table 36 presents the pooled treatment effect coefficients if each individual study were removed from the analysis (one at a time). For example, the combined effect for all 22 effect sizes is 0.109; if the Andreou study is dropped from the set of studies, the pooled estimate decreases to 0.088. Figure 8 depicts this analysis in a standard error bar chart.
Table 36 Influence of Individual Studies on Pooled Treatment Effect on Bullying Outcomes

<table>
<thead>
<tr>
<th>Study omitted</th>
<th>Coefficient</th>
<th>98% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>Andreou</td>
<td>.088</td>
<td>-.024</td>
</tr>
<tr>
<td>Baldry1</td>
<td>.127</td>
<td>.017</td>
</tr>
<tr>
<td>Baldry2</td>
<td>.095</td>
<td>-.015</td>
</tr>
<tr>
<td>Boulton</td>
<td>.111</td>
<td>.000</td>
</tr>
<tr>
<td>Cross</td>
<td>.116</td>
<td>.007</td>
</tr>
<tr>
<td>DeRosier</td>
<td>.110</td>
<td>-.004</td>
</tr>
<tr>
<td>Fekkes</td>
<td>.111</td>
<td>.001</td>
</tr>
<tr>
<td>Frey</td>
<td>.112</td>
<td>.001</td>
</tr>
<tr>
<td>Hunt</td>
<td>.116</td>
<td>.007</td>
</tr>
<tr>
<td>Jenson</td>
<td>.111</td>
<td>-.006</td>
</tr>
<tr>
<td>Melton</td>
<td>.107</td>
<td>-.002</td>
</tr>
<tr>
<td>Menard</td>
<td>.110</td>
<td>.000</td>
</tr>
<tr>
<td>Menesini</td>
<td>.099</td>
<td>-.013</td>
</tr>
<tr>
<td>Mikheev</td>
<td>.098</td>
<td>-.012</td>
</tr>
<tr>
<td>Rahey1</td>
<td>.110</td>
<td>.001</td>
</tr>
<tr>
<td>Rahey2</td>
<td>.110</td>
<td>.002</td>
</tr>
<tr>
<td>Rican</td>
<td>.105</td>
<td>-.004</td>
</tr>
<tr>
<td>Rosenbluth</td>
<td>.113</td>
<td>.003</td>
</tr>
<tr>
<td>Stevens1</td>
<td>.105</td>
<td>-.006</td>
</tr>
<tr>
<td>Stevens2</td>
<td>.130</td>
<td>.018</td>
</tr>
<tr>
<td>Taylor</td>
<td>.110</td>
<td>.001</td>
</tr>
<tr>
<td>Turpeau</td>
<td>.110</td>
<td>-.000</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>.109</strong></td>
<td><strong>.001</strong></td>
</tr>
</tbody>
</table>
Table 36 and Figure 8 demonstrate that despite the statistically significant overall pooled effect size of .109, the claim that prevention programs are significantly impact bullying outcomes is tenuous at best. Table 36 shows that if any of 10 studies are dropped from the analysis (Andreou, Baldry2, DeRosier, Jenson, Melton, Menesini, Mikheev, Rican, Stevens1, or Turpeau), the overall pooled estimate would drop to a non-statistically significant level. In other words, the influence analysis does not demonstrate that one or two studies are exerting an undue influence on the overall estimate of treatment effectiveness, but that the pooled effect is small enough that it is substantially affected by many of the effect size included in the analysis.
Meta-Analysis of Victimization Outcomes

Pooled Results

The analytic approach undertaken for the analysis of victimization outcomes is very similar to the approach applied to the set of effect sizes assessing bullying outcomes. The total $Q$-statistic for homogeneity of the victimization effect sizes was not statistically significant ($Q=22.859$ with 24 degrees of freedom, $p=.528$), suggesting a lack of heterogeneity among studies. Figure 9 presents a Galbraith plot of each effect size's $Z$-statistic against its inverse standard error. Two studies fall outside the upper bound of the 95% confidence interval, which suggests a small degree of heterogeneity in the set of victimization effect sizes. Nevertheless, the vast majority of the set of studies appears homogeneous, and the $Q$-statistic was non-significant. As such, the fixed effects model was used to estimate the pooled treatment effect for victimization outcomes.

Figure 9 Galbraith Plot of Study Heterogeneity for Victimization Meta-Analysis
The 25 victimization effect sizes, standard errors, variances, inverse variance weights, and relative weights are presented in Table 37. According to Cohen's conventions (1988), the majority of the effect sizes were small (n=21), with four effect sizes in the medium range.

Table 37 Study-Level Data for the Meta-Analysis of Victimization Outcomes

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Effect size (ES)</th>
<th>Standard error (SE)</th>
<th>Variance (v)</th>
<th>Inverse variance weight (w)</th>
<th>Relative weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alsaker (2001)</td>
<td>0.172</td>
<td>0.197</td>
<td>0.039</td>
<td>25.64</td>
<td>6.4%</td>
</tr>
<tr>
<td>Andreou (2007)</td>
<td>0.007</td>
<td>0.289</td>
<td>0.084</td>
<td>11.95</td>
<td>3.0%</td>
</tr>
<tr>
<td>Baldry #1 (2004)</td>
<td>0.628</td>
<td>0.274</td>
<td>0.075</td>
<td>13.33</td>
<td>3.3%</td>
</tr>
<tr>
<td>Baldry #2 (2004)</td>
<td>0.052</td>
<td>0.504</td>
<td>0.254</td>
<td>3.94</td>
<td>1.0%</td>
</tr>
<tr>
<td>Bauer (2007)</td>
<td>0.070</td>
<td>0.166</td>
<td>0.028</td>
<td>36.23</td>
<td>9.1%</td>
</tr>
<tr>
<td>Beran (2004)</td>
<td>0.060</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>3.5%</td>
</tr>
<tr>
<td>Cross (2004)</td>
<td>0.125</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>2.6%</td>
</tr>
<tr>
<td>DeRosier (2004a)</td>
<td>0.182</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>5.6%</td>
</tr>
<tr>
<td>Fekkes (2006b)</td>
<td>0.408</td>
<td>0.308</td>
<td>0.095</td>
<td>10.53</td>
<td>2.6%</td>
</tr>
<tr>
<td>Frey (2005)</td>
<td>0.174</td>
<td>0.470</td>
<td>0.221</td>
<td>4.52</td>
<td>1.1%</td>
</tr>
<tr>
<td>Hunt (2007)</td>
<td>-0.260</td>
<td>0.460</td>
<td>0.212</td>
<td>4.72</td>
<td>1.2%</td>
</tr>
<tr>
<td>Jenson (2007)</td>
<td>0.069</td>
<td>0.232</td>
<td>0.054</td>
<td>18.52</td>
<td>4.6%</td>
</tr>
<tr>
<td>Melton (1998)</td>
<td>-0.146</td>
<td>0.221</td>
<td>0.049</td>
<td>20.41</td>
<td>5.1%</td>
</tr>
<tr>
<td>Menard (2008)</td>
<td>0.150</td>
<td>0.359</td>
<td>0.129</td>
<td>7.75</td>
<td>1.9%</td>
</tr>
<tr>
<td>Menesini (2003)</td>
<td>0.148</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>2.6%</td>
</tr>
<tr>
<td>Mikheev (2005)</td>
<td>0.671</td>
<td>0.200</td>
<td>0.040</td>
<td>25.00</td>
<td>6.3%</td>
</tr>
<tr>
<td>Rahey #1 (2002)</td>
<td>0.658</td>
<td>0.231</td>
<td>0.053</td>
<td>18.74</td>
<td>4.7%</td>
</tr>
<tr>
<td>Rahey #2 (2002)</td>
<td>0.017</td>
<td>0.378</td>
<td>0.143</td>
<td>7.00</td>
<td>1.8%</td>
</tr>
<tr>
<td>Rican (1996)</td>
<td>0.055</td>
<td>0.222</td>
<td>0.049</td>
<td>20.29</td>
<td>5.1%</td>
</tr>
<tr>
<td>Rosenbluth (2004)</td>
<td>0.217</td>
<td>0.204</td>
<td>0.042</td>
<td>24.03</td>
<td>6.0%</td>
</tr>
<tr>
<td>Stevens #1 (2000)</td>
<td>0.031</td>
<td>0.402</td>
<td>0.162</td>
<td>6.19</td>
<td>1.5%</td>
</tr>
<tr>
<td>Stevens #2 (2000)</td>
<td>0.533</td>
<td>0.538</td>
<td>0.289</td>
<td>3.45</td>
<td>0.9%</td>
</tr>
<tr>
<td>Taylor (2006)</td>
<td>-0.215</td>
<td>0.240</td>
<td>0.058</td>
<td>17.36</td>
<td>4.3%</td>
</tr>
<tr>
<td>Turpeau (1998)</td>
<td>0.297</td>
<td>0.142</td>
<td>0.020</td>
<td>49.75</td>
<td>12.5%</td>
</tr>
<tr>
<td>Woods (2007)</td>
<td>0.002</td>
<td>0.272</td>
<td>0.074</td>
<td>13.51</td>
<td>3.4%</td>
</tr>
<tr>
<td><strong>Σ</strong></td>
<td><strong>4.105</strong></td>
<td><strong>7.416</strong></td>
<td><strong>2.481</strong></td>
<td><strong>399.580</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

---

108 Portions of the data for some of the included studies were obtained from Ttofi, Farrington, and Baldry (2008), as they were unavailable in the published reports themselves. Specifically, standard deviations were obtained for the study by Baldry and Farrington (2004), means and standard deviations were obtained for the study by Rahey and Craig (2002), and means and standard deviations were obtained for the study by Hunt (2007).
Table 38 presents the weighted mean effect size, 95% confidence interval, and Z-score for the set of victimization effect sizes. The pooled effect size is 0.188 (95% C.I. 0.090, 0.286), with a Z-test value of 3.763 (p< .001). In other words, the mean effect of bullying prevention programs on victimization outcomes is statistically significant.

### Table 38 Fixed Effects Meta-Analysis for Victimization Outcomes

<table>
<thead>
<tr>
<th>Pooled estimate</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>Z</th>
<th>p-value</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.188</td>
<td>0.090</td>
<td>0.286</td>
<td>3.763</td>
<td>&lt;0.001</td>
<td>25</td>
</tr>
</tbody>
</table>

Test for heterogeneity: Q= 22.859 on 24 degrees of freedom (p= 0.528), $I^2 = 0\%$

A forest plot of the effect size estimates and confidence intervals is depicted in Figure 10. The diamond at the bottom of the figure represents the pooled estimate of the effect size (0.188) with its 95% confidence intervals indicated by the left and right edges of the diamond. The confidence interval for the pooled effect does not cross zero, clearly demonstrating a significant and positive treatment effect.
Investigating Between-Study Heterogeneity

Potential predictors of treatment impacts on victimization outcomes were explored through two modeling approaches: subgroup analysis and linear meta-regression. The subgroup analysis approach was described earlier and is identical to the analytic strategy implemented for the assessment of bullying outcomes.

Table 39 presents results from the subgroup analyses of five study characteristics on bullying outcomes. The bivariate subgroup estimates were compared using the analog to the ANOVA method to assess whether the categorical variable groupings account for any of the variability between effect sizes. None of the $Q_b$'s for the study characteristic variables were statistically significant; suggesting that publication type, geographic location, research design, scale of outcome measure, and sample size are not significant factors in predicting the magnitude of the treatment impact on victimization outcomes.
Table 39 Subgroup Analysis of Study Characteristics on Victimization Outcomes

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Effect size and Q-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Publication type</strong></td>
<td></td>
</tr>
<tr>
<td>Peer-reviewed (n=20)</td>
<td>$ES = 0.193$</td>
</tr>
<tr>
<td>Non peer-reviewed (n=5)</td>
<td>$ES = 0.149$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.08 \sim \chi^2_{1}$, $p=.777$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 22.78 \sim \chi^2_{23}$, $p=.474$</td>
</tr>
<tr>
<td><strong>2 Geographic location</strong></td>
<td></td>
</tr>
<tr>
<td>North America (n=14)</td>
<td>$ES = 0.181$</td>
</tr>
<tr>
<td>Europe or Australia (n=11)</td>
<td>$ES = 0.196$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.21 \sim \chi^2_{1}$, $p=.647$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 22.84 \sim \chi^2_{23}$, $p=.470$</td>
</tr>
<tr>
<td><strong>3 Research design</strong></td>
<td></td>
</tr>
<tr>
<td>Randomized (n=11)</td>
<td>$ES = 0.122$</td>
</tr>
<tr>
<td>Non-randomized (n=14)</td>
<td>$ES = 0.281$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 2.43 \sim \chi^2_{1}$, $p=.119$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 20.43 \sim \chi^2_{23}$, $p=.616$</td>
</tr>
<tr>
<td><strong>4 Scale of outcome measure</strong></td>
<td></td>
</tr>
<tr>
<td>Dichotomous (n=8)</td>
<td>$ES = 0.217$</td>
</tr>
<tr>
<td>Continuous (n=17)</td>
<td>$ES = 0.172$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.19 \sim \chi^2_{1}$, $p=.663$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 22.67 \sim \chi^2_{23}$, $p=.480$</td>
</tr>
<tr>
<td><strong>5 Sample size</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;400 students (n=13)</td>
<td>$ES = 0.224$</td>
</tr>
<tr>
<td>400+ students (n=12)</td>
<td>$ES = 0.163$</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>$Q_B = 0.36 \sim \chi^2_{1}$, $p=.549$</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>$Q_W = 22.50 \sim \chi^2_{23}$, $p=.490$</td>
</tr>
</tbody>
</table>

* Statistically significant at $p< .01$

Table 40 presents results from the subgroup analyses of treatment program elements on victimization outcomes. Significant measures of between-groups heterogeneity were found for four of the treatment elements: level of implementation, whole school anti-bullying policy, individual work with bullies or victims, and peer mediation/mentoring. This suggests that the mean effect sizes produced by each subgroup are significantly different from one another, and differ by more than would be expected as a result of sampling error. More specifically, programs implemented at the classroom level, as opposed to the whole school level, were
related to stronger reductions in levels of victimization. In addition, programs that did not incorporate a whole school anti-bullying policy, did not use individualized work with students identified as bullies or victims, and did not involve peer mediation/mentoring were related to greater treatment success in reducing levels of victimization. The remaining six program elements measured in the analysis were not significant factors in predicting the magnitude of treatment effects on victimization outcomes.

Table 40 Subgroup Analysis of Treatment Elements of Victimization Outcomes

<table>
<thead>
<tr>
<th>Treatment element</th>
<th>Effect size and Q-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Level of implementation</strong></td>
<td></td>
</tr>
<tr>
<td>Whole school (n=13)</td>
<td>ES = 0.019</td>
</tr>
<tr>
<td>Classroom (n=11)</td>
<td>ES = 0.350</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 10.46 \sim \chi^2_{1,} ), ( p=.001^* )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 12.40 \sim \chi^2_{22,} ), ( p=.949 )</td>
</tr>
<tr>
<td><strong>2 Program duration</strong></td>
<td></td>
</tr>
<tr>
<td>( \leq 12 ) weeks (n=11)</td>
<td>ES = 0.113</td>
</tr>
<tr>
<td>&gt;12 weeks (n=14)</td>
<td>ES = 0.229</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 1.23 \sim \chi^2_{1,} ), ( p=.267 )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 21.63 \sim \chi^2_{23,} ), ( p=.543 )</td>
</tr>
<tr>
<td><strong>3 Mean participant age</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;11 years (n=11)</td>
<td>ES = 0.131</td>
</tr>
<tr>
<td>11+ years (n=14)</td>
<td>ES = 0.218</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 0.69 \sim \chi^2_{1,} ), ( p=.406 )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 22.17 \sim \chi^2_{23,} ), ( p=.510 )</td>
</tr>
<tr>
<td><strong>4 Curriculum</strong></td>
<td></td>
</tr>
<tr>
<td>Used curriculum (n=21)</td>
<td>ES = 0.163</td>
</tr>
<tr>
<td>No curriculum (n=4)</td>
<td>ES = 0.364</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 1.79 \sim \chi^2_{1,} ), ( p=.181 )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 21.07 \sim \chi^2_{23,} ), ( p=.577 )</td>
</tr>
<tr>
<td><strong>5 Whole-school anti-bullying policy</strong></td>
<td></td>
</tr>
<tr>
<td>Whole-school policy (n=11)</td>
<td>ES = 0.023</td>
</tr>
<tr>
<td>No whole-school policy (n=14)</td>
<td>ES = 0.294</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 6.97 \sim \chi^2_{1,} ), ( p=.008^* )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 15.89 \sim \chi^2_{23,} ), ( p=.860 )</td>
</tr>
<tr>
<td><strong>6 Classroom anti-bullying rules</strong></td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Classroom rules (n=13)</td>
<td>( ES = 0.177 )</td>
</tr>
<tr>
<td>No classroom rules (n=12)</td>
<td>( ES = 0.202 )</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
<td>( Q_B = 0.07 \sim \chi^2_1, \ p=.791 )</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
<td>( Q_W = 22.79 \sim \chi^2_{23}, \ p=.473 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>7 Teacher training</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher training (n=19)</td>
</tr>
<tr>
<td>No teacher training (n=6)</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>8 Individual work with bullies and/or victims</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual bully/victim work (n=11)</td>
</tr>
<tr>
<td>No individual bully/victim work (n=14)</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>9 Peer mediation or mentoring</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer mediation or mentoring (n=8)</td>
</tr>
<tr>
<td>No peer mediation or mentoring (n=17)</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>10 Parent information or meetings</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents involved (n=13)</td>
</tr>
<tr>
<td>No parents involved (n=12)</td>
</tr>
<tr>
<td>Between-studies heterogeneity</td>
</tr>
<tr>
<td>Within-studies heterogeneity</td>
</tr>
</tbody>
</table>

*Statistically significant at \( p< .05 \)

Given the significant results in four of the subgroup analyses, heterogeneity was further explored using a modified weighted least squares regression analysis.\(^{109}\) The regression model represents an independent assessment of the moderator effects, to examine whether a relationship exists above and beyond what was found in the subgroup analyses. This analysis also addresses the issue of possible non-independence of the moderator variables.

\(^{109}\) A mixed effects model was implemented, which estimates both fixed and random effects variance components. The random effects variance component is estimated using a method of moments estimator (a non-iterative procedure). The random effects variance component can also be estimated using an iterative procedure based on restricted maximum likelihood, maximum likelihood, or empirical Bayes methods (Sharp, 1998).
To select variables for the model, all of the 10 binary variables used in the subgroup moderator analyses of treatment elements were assessed for inter-correlations. Eight of the bivariate correlations were statistically significant; with five characterized as large 
($0.50 < |r| < 1.00$), and three considered moderate ($0.30 < |r| < 0.50$). Due to the limited number of degrees of freedom associated with having just 25 cases, only six of the variables were included in the meta-regression model. These include the level of program implementation, individual work with bullies or victims, whole-school anti-bullying policy, peer mediation/mentoring, parent information, and teacher training variables. The first four variables were selected based on their significant $Q_b$ statistics found in the subgroup analysis, while the parent information and teacher training variables were selected because they were strongly correlated with one or more of these four variables. Table 41 presents results from the meta-regression on victimization outcomes. The coefficient of -.49 for the whole-school anti-bullying policy was significant and negative ($Z = -2.53$, $p = .011$), while the coefficient for the parent information/meetings variable was significant and positive ($\beta = .46$, $Z = 2.31$, $p = .02$). These results suggest that programs without whole-school policies, and programs with parent involvement, are related to stronger treatment impacts on victimization. Interestingly, the level of implementation (whole school or classroom), individual work with bullies or victims variables, and peer mediation/mentoring, all three of which had been significant in the moderator analysis, were not significant in the meta-regression.

| Regressor                                | Coef. | Std. err. | $Z (p>|Z|)$ | 95% Conf. Interval |
|------------------------------------------|-------|-----------|------------|--------------------|
| Level of implementation                  | -.066 | .099      | -0.67 (0.505) | -.261 -.129        |
| Individual work w/bullies/victims       | -.067 | .230      | -0.29 (0.770) | -.519 .384         |
| Whole-school anti-bullying policy        | -.494 | .195      | -2.53 (0.011)* | -.876 -.112        |
| Peer mediation/mentoring                 | -.065 | .211      | -0.31 (0.759) | -.478 .348         |
| Parent information/meetings              | .460  | .199      | 2.31 (0.021)* | .070 .850          |
| Teacher training                         | -.126 | .165      | -0.76 (0.446) | -.449 .198         |
| Constant                                 | .332  | .106      | 3.14 (0.002)* | .125 .539          |
Sensitivity Analysis

Publication bias

Table 42 and Figures 11 and 12 present an analysis of publication bias for the studies assessing victimization outcomes. Neither the Begg and Mazumdar test for the correlation between effect size estimates and their variances nor the Egger test for funnel plot asymmetry were statistically significant, which suggests a lack of publication bias.

Table 42 Tests for Publication Bias in Victimization Outcomes Data Set

<table>
<thead>
<tr>
<th>Begg &amp; Mazumdar's Test</th>
<th>Adj. Kendall's Score</th>
<th>Std. dev.</th>
<th>No. of studies</th>
<th>Z (p-value)</th>
<th>Z (p-value) (continuity corrected)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-11</td>
<td>42.81</td>
<td>25</td>
<td>-0.26(0.797)</td>
<td>0.23(0.815)</td>
</tr>
</tbody>
</table>

| Egger's Test | Std effect | Coefficient | Std. error | t (p>|t|) | 95% Conf. Interval |
|--------------|------------|-------------|------------|---------|-------------------|
|              | Slope      | .298        | .156       | 1.91 (0.068) | -0.024      0.620 |
|              | bias       | -.463       | .622       | -0.74 (0.464) | -1.750      0.825 |
Figure 11 Begg's Funnel Plot of Study Effect Size vs. Standard Error, Victimization Outcomes
The sloping line in the Egger plot depicts the regression of the standardized effect estimates on precision, with results indicating that the intercept does not deviate significantly from zero ($t$-statistic for bias = -.74, $p = .464$).

**Solution sensitivity**

Begg's funnel plot (Figure 11) was examined for whether any effect sizes appeared unusual. Two studies were shown to fall outside of the pseudo-95% confidence intervals, and were identified as the studies by Alsaker and Valkanover (2001) ($ES=.658$, $SE=.231$), and Woods et al. (2007) ($ES=.671$, $SE=.200$). Alsaker's study had sample sizes of 152 treatment and 167 control, with a relative weight of 4.7 percent. The study by Woods et al. had sample sizes of 284 treatment and 116 control, with a relative weight of 6.3 percent. Although these studies had relatively small standard errors associated with their relatively large effect sizes, due to their rather small sample sizes they did not exert a disproportionate influence on the pooled treatment effect estimate. As such, the studies were left unaltered in the set of observations.
An examination of solution sensitivity to the individual effect size estimates is presented in Table 43 and Figure 13. Table 43 displays the pooled treatment effect coefficients for the victimization outcomes if each individual study were removed from the analysis one at a time; Figure 13 depicts this analysis in a standard error bar chart.

### Table 43 Influence of Individual Studies on Pooled Treatment Effect on Victimization Outcomes

<table>
<thead>
<tr>
<th>Study omitted</th>
<th>Coefficient</th>
<th>98% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Alsaker</td>
<td>.165</td>
<td>.265</td>
</tr>
<tr>
<td>Andreou</td>
<td>.189</td>
<td>.291</td>
</tr>
<tr>
<td>Baldry1</td>
<td>.194</td>
<td>.293</td>
</tr>
<tr>
<td>Baldry2</td>
<td>.173</td>
<td>.273</td>
</tr>
<tr>
<td>Bauer</td>
<td>.191</td>
<td>.290</td>
</tr>
<tr>
<td>Beran</td>
<td>.190</td>
<td>.288</td>
</tr>
<tr>
<td>Cross</td>
<td>.195</td>
<td>.296</td>
</tr>
<tr>
<td>DeRosier</td>
<td>.200</td>
<td>.303</td>
</tr>
<tr>
<td>Fekkes</td>
<td>.186</td>
<td>.287</td>
</tr>
<tr>
<td>Frey</td>
<td>.193</td>
<td>.293</td>
</tr>
<tr>
<td>Hunt</td>
<td>.190</td>
<td>.289</td>
</tr>
<tr>
<td>Jenson</td>
<td>.173</td>
<td>.277</td>
</tr>
<tr>
<td>Melton</td>
<td>.191</td>
<td>.289</td>
</tr>
<tr>
<td>Menard</td>
<td>.195</td>
<td>.294</td>
</tr>
<tr>
<td>Menesini</td>
<td>.189</td>
<td>.289</td>
</tr>
<tr>
<td>Mikheev</td>
<td>.182</td>
<td>.282</td>
</tr>
<tr>
<td>Rahey1</td>
<td>.188</td>
<td>.287</td>
</tr>
<tr>
<td>Rahey2</td>
<td>.194</td>
<td>.292</td>
</tr>
<tr>
<td>Rican</td>
<td>.185</td>
<td>.284</td>
</tr>
<tr>
<td>Rosenbluth</td>
<td>.206</td>
<td>.307</td>
</tr>
<tr>
<td>Stevens1</td>
<td>.194</td>
<td>.294</td>
</tr>
<tr>
<td>Stevens2</td>
<td>.206</td>
<td>.307</td>
</tr>
<tr>
<td>Taylor</td>
<td>.189</td>
<td>.288</td>
</tr>
<tr>
<td>Turpeau</td>
<td>.189</td>
<td>.289</td>
</tr>
<tr>
<td>Woods</td>
<td>.156</td>
<td>.257</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>.188</strong></td>
<td><strong>.286</strong></td>
</tr>
</tbody>
</table>
No extreme effect size values or standard errors were observed in the set of included studies, indicating that no study was exerting an undue influence on the overall pooled mean effect size. Furthermore, the overall pooled effect size of .188 is sufficiently robust so as not to be significantly affected by the removal of any one study in the analysis.

DISCUSSION

Results from this analysis suggest that as a whole, bullying prevention programs are successful at reducing the level of victimization in schools, but have questionable effectiveness at reducing the level of bullying. After participating in school-based bullying prevention programs, students report effect size of .188 for a reduction in victimization, and an effect size of .109 for reduction in bullying others. While both effects are statistically significant, the
pooled effect for bullying outcomes is so small as to be highly contingent on the inclusion of specific studies in the analysis. These results are reasonably consistent with the results reported in the previous meta-analysis by Ferguson et al. (2007), who found very small, positive effects for bullying outcomes that were not statistically significant (Ferguson did not assess victimization outcomes). These results diverge in part from the results in the meta-analysis by Ttofi and Farrington (2009), who found significant effects for both bullying and victimization, with a substantially larger pooled effect size for victimization than found in the current analysis. Ttofi and Farrington also conducted subgroup analyses of potential moderators of treatment impact, and found a range of variables to be significant predictors of between-group heterogeneity. Some of these moderator variables overlapped with the variables used in the current analysis. For example, with respect to victimization outcomes the authors found that peer mentoring or mediation, parent training, program duration (longer), age of participants (older), geographic location (Europe), and non-randomized designs were associated with stronger program impacts. For the most part these findings were not replicated in the current study; of the moderator variables common to both analyses, only non-randomized studies, level of program implementation (classroom), and individual work with bullies or victims demonstrated significant between-groups heterogeneity and a link with stronger treatment impacts. Ttofi and Farrington used somewhat different inclusion criteria than the current analysis (see the Introduction of the current paper for more details on how the studies differed), which likely accounts for some of the disparity in treatment outcomes. Furthermore, the authors used different methods to calculate effect sizes, which also are likely to have had an impact on study conclusions (see the technical appendix in Ttofi, Farrington, & Baldry, 2008, for more details on effect size calculations).

Despite non-significant overall $Q$-statistics for total homogeneity in the sets of effect sizes addressing bullying and victimization outcomes, the current study explored study characteristics and treatment program characteristics in a subgroup moderator analysis to determine whether any such characteristics might be related to stronger treatment impacts. No significant between-groups factors were identified for the set of bullying moderator variables,
but four of the victimization moderators appeared as significant in explaining between-study heterogeneity. Significant effects were found for treatments implemented at the classroom level (versus at the whole school level), treatments that did not involve whole school anti-bullying policies, treatments that did not incorporate individualized work with bullies or victims, and treatments that did not use peer mediation/mentoring. In other words, classroom-level programs and the absence of other three treatment characteristics were associated with lower rates of victimization. Using a mixed-effects meta-regression to further disentangle the effects of study and treatment characteristics on victimization outcomes, results showed that programs using parent information or meetings, and programs not using whole-school anti-bullying policies were associated with stronger treatment impacts. The negative finding for the whole-school anti-bullying policies is counterintuitive and should be explored in further research. Various explanations could be offered to explain this result: for example, perhaps students view whole-school policies as draconian and seek to rebel against stated policies as a way to express autonomy, or perhaps students at schools with formal anti-bullying policies are more adept at identifying their involvement in bullying situations (resulting in higher reported rates of victimization). In any case, these results suggest that bullying prevention programs may be more successful at reducing victimization if they include parent involvement through informational meetings and/or newsletters/guides, but do not include individualized work with bullies and victims, peer mediation/mentoring, or formal school policies against bullying.

The overall finding that programs are effective at reducing victimization but are not particularly effective at reducing bullying is intriguing, and highlights the importance of considering the outcomes from bullying prevention programs as being two-pronged, with potential impacts for bullies and victims not necessarily analogous. How can it be that programs are successful at reducing victimization but are not as successful at simultaneously reducing bullying? This finding warrants further research and consideration of how to improve programs to better target bully perpetrators. Several preliminary hypotheses may explain this puzzling outcome. For example, it is possible that while anti-bullying programs help some students gain skills to avoid being victimized in the future, other students may not be as
successful and as a result end up the unfortunate subjects of even more bullying (i.e., since previous victims have removed themselves from the pool of potential targets for bullies, yet the rate of bully perpetration has not decreased). This hypothesis would be consistent with Perry, Williard, and Perry (1990), who surmise:

*It seems likely that… aggressive children "try out" aggressive responses toward a variety of targets, especially those with qualities predictive of reinforcement of aggression (e.g., slightness of build, manifest anxiety). Then, as aggressors learn the reactions of victims and cognitively represent these reactions in the form of outcome expectancies and values, the field of victims becomes increasingly narrowed.* (p. 1322)

It may also be an issue of measurement: perhaps the anti-bullying programs are in fact not successful at reducing rates of actual victimization; instead, throughout the course of the program victims come to realize that behavior they classified as "being bullied" at pretest does not in fact qualify as true bullying behavior. As such, a reduction in victimization rates on the post-test may be a reflection of measurement error at pretest. The reverse is also true: perhaps a true decline in bullying is produced by the anti-bullying programs, but at pretest some students did not realize that their actions constitute bullying behaviors. If so, rates of bullying at pretest might be artificially low, and any resulting decrease in rates of bullying might not be detected by the measure. Also relating to measurement might be a change in students' willingness to classify themselves as victims. After completing an anti-bullying intervention, participants may feel increased shame in admitting they are still the victims of bullying, and therefore may underestimate their levels of victimization on the posttest surveys. Other potential explanations exist, and the results of the current study do not allow for a comprehensive examination of the reason for the discrepancy seen in treatment effectiveness for bullies versus victims.

**Limitations**

In order to draw conclusions about the effectiveness of bullying prevention programs in schools, the current study implemented systematic and rigorous methods to identify and
select studies for inclusion, extract and transform data for analysis, and conduct reasonably comprehensive meta-analyses of the data. Nevertheless, some limitations to this research are apparent and are discussed below.

Although every attempt was made to identify and retrieve all bullying prevention program evaluations that met the inclusion criteria protocol, the possibility remains that not all such evaluations were included in the analyses. Despite the use of 14 electronic databases across several fields of study, it may be that certain studies are not referenced in these databases, or are not cataloged by the subject terms used in this study's search strategy. The complementary method of reviewing the bibliographies of all included studies and prior meta-analyses and systematic reviews helps to counter the possibility that key studies were missed, but is not a guarantee. Furthermore, although neither the bullying nor the victimization assessments of publication bias suggested reason for concern, it remains possible that some evaluations of bullying prevention programs have been conducted yet lie unpublished due to non-significant results.

It is also possible that the reasonably strict inclusion criteria, in particular the requirements for a comparison group along with either a pretest assessment or a randomized study design, limited the set of included studies in such as way as to affect the summative conclusions of the analyses. Without question, the set of included studies would have been substantially larger without one or both of these requirements.\textsuperscript{110} For example, a large number of studies implemented a one-group pretest-posttest design, a two-group posttest-only design without random assignment, or a single group posttest design to study the effects of a bullying prevention program.\textsuperscript{111} The addition of these studies to the set of observations included in the analysis would likely have had a large impact on the meta-analytic findings. However, a larger set of studies is not necessarily desirable; it is possible that the lower level of rigor in these

\textsuperscript{110} For example, see the set of included studies in the meta-analysis by Ttofi and Farrington (2009), which did not implement such stringent inclusion requirements.

\textsuperscript{111} It should be noted that the well-known Olweus Bullying Prevention Program studies were excluded as a result of these criteria. Olweus (1994) used a time-lagged cohort design to evaluate a bullying prevention program in a sample of 2,500 Norwegian students from 1983-1985. Often referred to as the Bergen Study, and frequently mentioned in discussions of effective bullying prevention, this design was not considered rigorous enough to meet the current inclusion criteria protocol.
studies might lead to spurious outcomes. See Cook and Campbell (1979) for a detailed review of internal validity concerns with respect to quasi-experimental and non-experimental designs. This study exercised a conservative approach, focusing on a small pool of the most rigorous studies in lieu of a larger set of studies with less convincing research designs.

The comparability of the treatment programs in the included set of studies is also a concern, as the most frequent criticism of meta-analysis is the pooling together of incommensurate studies. It is clear that the bullying prevention programs used in the current analyses were varied with respect to program length, age of participants, and treatment elements such as classroom curriculum, whole school anti-bullying policies, parent involvement, and so on. The extent to which these program differences could be characterized as 'mixing apples with oranges' is debatable. However, if a summative conclusion about the effects of school-based bullying prevention programs is desired – no matter what the nature of such programs, then the current set of included studies seems reasonable.

Even with the strict inclusion criteria, other issues complicating interpretation of the results from the included studies exist. To the extent that some of the studies had methodological weaknesses, their results may be biased and subsequently the results of the current analysis may be biased as well. For instance, a lack of treatment implementation fidelity may have been present, which in essence means that the evaluation is assessing a different program than the one described in the research report. As an example, a bullying prevention curriculum may have been designed with 10 sessions, but in practice the classroom teacher may have implemented only 8 sessions, and the attendance rate at the 8 sessions may have been below 50 percent. In this hypothetical situation, any evaluation of this program would clearly not be assessing the impact of the 10-session program as originally designed. To the extent possible, methodological design flaws that were hypothesized to have a potential biasing impact on study outcomes were used to eliminate such studies from the included set (e.g., if attrition in the experimental group was substantially larger than in the control group, to the point where the comparability of the treatment groups was reduced to an unsatisfactory level,
the study was excluded). However, as is always the case, only observable and reported study design and implementation flaws were accounted for.

Another example of a potential study-level weakness is the possibility of testing effects. As discussed previously, it is possible that respondents' sensitivity to bullying behaviors was increased through exposure to the pretest or participation in the program itself, and this increased sensitivity may be linked to higher (or lower) rates of reporting at posttest. For example, prior to participation in the program students may not have realized that certain forms of relational or verbal aggression qualify as forms of bullying, such as spreading rumors or teasing. In this case, despite no change in actual victimization rates, following participation in the program students might report a higher frequency of victimization than what they reported at pretest, owing to a change in their definition of what constitutes "bullying." If true, this would suggest that the rates of victimization at pretest were artificially low, and an analysis assessing change scores between pretest and posttest would be misleading. The extent to which this type of problem was a concern for the 23 studies included in the current analysis is unknown.

Conclusion

The sum of the literature on bullying prevention programs provides evidence that they are effective at reducing the occurrence of victimization in schools, but provides minimal evidence of their effectiveness at reducing bullying. The significant finding for a reduction in victimization is encouraging, and suggests that programs are somewhat successful in tackling the problem of bullying in schools. The weak effect for bullying outcomes is disappointing, and is somewhat unexpected given the considerable amount of effort that has been devoted to understanding and preventing bullying in schools. The fact that bullying has been identified as a cross-cultural and prevalent problem worldwide suggests that the behavior may stem from an innate human motivation to aggress and dominate in peer groups. Therefore, bullying behaviors may be particularly difficult to change, and prevention programs may face major challenges in producing even small positive effects.
If bullying were an innocuous activity that had not previously been linked to negative outcomes for both those involved in perpetration and those subjected to victimization, then the observed reduction in victimization rates might be viewed as evidence of overall success. Results from previous research show that this is clearly not the case (see Chapter 1 of the dissertation), and children who avoid both bullying and victimization are less likely to experience negative outcomes than those who end up as willing or unwilling participants in bullying behaviors. Knowing that reducing bullying behavior will serve to decrease negative outcomes such as violence, school failure, or depression is not particularly useful if we are unable to determine how to effectively reduce this behavior. While a reduction in rates of victimization is clearly important and is gratifying to observe in a pooled estimate across 25 program evaluations, it is inefficient to implement programs that are not effective at reducing or eliminating bullying as well. This research suggests that, as a whole, current prevention efforts are not adequately achieving their goals of reducing the problem of bullying in schools, since they do not appear to be sufficiently effective for bully perpetrators. This information is valuable because it precludes any inclination for complacency with respect to developing new programs. Clearly, much more work needs to be done. Future studies should be directed towards determining why programs might be more successful with victims of bullying rather than with perpetrators. Research should also seek to identify existing program components or develop novel program components that are more likely to bring about successful reductions in bullying as well as victimization. Naturally, it is crucial that these programs are subject to rigorous evaluation, so that schools are provided with the requisite data to make evidence-based decisions with respect to bullying prevention.
References


---

112 Studies in bold font were included in Chapter 3's meta-analyses of program effects.


C.R.S. 22-32-109 (2001)


Reichardt, C. S., Minton, B. A., & Schellenger, J. D. (1980). The analysis of covariance (ANCOVA) and the assessment of treatment effects. In *Prevention evaluation research monograph II—Outcome* (Section D4, chap. VIII). Denver, CO: University of Denver, Department of Psychology.


320


Valentine, J. (2008 May). Fixed effects and homogeneity evaluation. Presentation at the annual meeting of the Campbell Collaboration, Vancouver, B.C.


Appendix 1
Index/Scale Variables Used for Matching

1. Physical Environment Index
   - This index consists of 5 survey items, including two self-reports from the youth questionnaire and 3 items reported by the interviewer. Each item was coded into risk categories, with scores of 0 to 2. Responses were summed, with an index score ranging from 0 to 7 – higher scores indicate a higher risk physical environment in the home.
   - Scores were calculated for respondents who answered at least four of the five scale items. Respondents who answered only four items were assigned a weighted score based on the 7-point scale (i.e., raw score*(5/5-missing)) (Moore et al., 1999).
   - The two self-report questions were asked only of youths aged 12-14 at baseline. They are as follows:
     1. In the past month, has your home usually had electricity and heat when you needed it?
     2. In a typical week, how many days from 0 to 7 do you hear gunshots in your neighborhood?
   - The items reported by interviewers are as follows:
     3. When you went to the respondent’s neighborhood/home, did you feel concerned for your safety?
     4. How well kept are most of the buildings on the street where the adult/youth resident lives?
     5. How well kept is the interior of the home in which the youth respondent lives?

2. Enriching Environment Index
   - This index consists of 3 dichotomous yes/no questionnaire items. Responses were summed, with a score ranging from 0 to 3. Higher scores indicate a more enriching home environment (Moore et al., 1999).
   - The questions were asked only of youths aged 12-14 at baseline. The survey items are as follows:
     1. In the past month, has your home usually had a computer?
     2. In the past month, has your home usually had a dictionary?
3. In a typical [school week/work week/week], did you spend any time taking extra classes or lessons for example, music, dance, or foreign language lessons?

3. Family Routines Index
   - The Family Routines index consists of 4 questionnaire items, measured on an 8-point scale ranging from 0 ("no days") to 7 ("all 7 days"). Responses were summed, with a score ranging from 0 to 28; higher scores indicate more days spent in routine activities with the family.
   - An index score was obtained for respondents who answered at least three out of four items; respondents who answered only three items were assigned a weighted score based on the 28-point scale (i.e., raw score*(4/4-missing)) (Moore et al., 1999).
   - The questions were asked only of youths aged 12-14 at baseline. Questionnaire items are as follows:
     1. In a typical week, how many days from 0 to 7 do you eat dinner with your family?
     2. In a typical week, how many days from 0 to 7 does housework get done when it is supposed to, for example cleaning up after dinner, doing dishes, or taking out the trash?
     3. In a typical week, how many days from 0 to 7 do you do something fun as a family such as play a game, go to a sporting event, go swimming and so forth?
     4. In a typical week, how many days from 0 to 7 do you do something religious as a family such as go to church, pray or read the scriptures together?

4. Parent Religiosity Scale
   - This scale consists of 6 dichotomous questionnaire items. Responses were summed, with a resulting score ranging from 0 to 6; higher scores indicate a higher level religiosity of the responding parent.
   - A score was obtained only for respondents who answered at least five out of the six questionnaire items; respondents who answered five items were assigned a weighted score based on the 6-point scale (i.e., raw score*(6/6-missing)) (Moore et al., 1999).
The Cronbach's alpha for this scale was .60, considered moderate in terms of internal consistency/reliability.

The questions were asked only of youths aged 12-16 at baseline, with questionnaire items as follows:

1. *I don't need religion to have good values.* (reverse coded)
2. *The Bible/Koran/Torah/religious teachings should be obeyed exactly as written in every situation.*
3. *I often ask God to help me make decisions.*
4. *God has nothing to do with what happens to me personally.* (reverse coded)
5. *I pray MORE than once a day.*
6. *In the past 12 months, how often have you attended a worship service (like church or synagogue service or mass)?* (1=once a month or more; 0=less than once a month)

5. **Substance Use Index**

- The Substance Use Index is calculated as the sum of 3 dichotomous questionnaire items, with a score ranging from 0 to 3. Higher scores indicate more experimentation with illegal substances (Moore et al., 1999).
- The questions were asked only of youths aged 12-16 at baseline. The NLSY97 survey items are as follows:
  1. *Have you ever smoked a cigarette?*
  2. *Have you ever had a drink of an alcoholic beverage? (By a drink we mean a can or bottle of beer, a glass of wine, a mixed drink, or a shot of liquor. Do not include childhood sips that you might have had from an older person's drink.)*
  3. *Have you ever used marijuana, for example: grass or pot, in your lifetime?*
6. Behavioral / Emotional Problems Scale

- The items measuring behavioral and emotional problems were adapted from the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1987), a standardized questionnaire of parent ratings of their children's problems and competencies.
- The scale included four questionnaire items that were measured on a 3-point scale, ranging from 0 ("not true") to 2 ("often true"). Responses were summed, with a total score ranging from 0 to 8; higher scores indicate more frequent and/or numerous behavior problems.
- Scores were obtained for respondents who answered at least three of the four items; respondents who answered only three of the items were assigned a weighted score based on the 8-point scale (i.e., rawscore*(4/4-missing) (Moore et al., 1999).
- Separate scales were created for boys and girls, with two identical items and two differing items. Cronbach's alpha was .53 for the girls' scale, and .51 for the boys' scale, which are considered somewhat low, yet adequate in terms of internal consistency/reliability.
- The questions were asked only of youths aged 12-14 at baseline. Questions were prefaced by the following statement:
  - The next few items describe youth. For each item, please indicate how well it describes you during the past 6 months. Is it not true, sometimes true, or often true of you? Please answer all items as well as you can even if some do not seem to apply to you.
- The items for girls:
  1. Your school work is poor.
  2. You have trouble sleeping.
  3. You lie or cheat.
  4. You are unhappy, sad, or depressed.
- The items for boys:
  1. You have trouble concentrating or paying attention.
  2. You don't get along with other kids.
3. You lie or cheat.
4. You are unhappy, sad, or depressed.
## Appendix 2
### Study and Treatment Characteristics (Study-Level)

#### Table 44 Key Study Characteristics

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Outcome</th>
<th>Pub. type</th>
<th>Location</th>
<th>Research design</th>
<th>Level of implementation</th>
<th>Prog. length (wks)</th>
<th>Student age in years (average age)</th>
<th>Out come source</th>
<th>Sample sizea,b (# students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Boulton (1996)</td>
<td>bully</td>
<td>journal article</td>
<td>Staffordshir e, England</td>
<td>random, simple</td>
<td>class</td>
<td>0.2</td>
<td>11-15 (13)</td>
<td>self-report</td>
<td>170</td>
</tr>
<tr>
<td>8. DeRosier (2004a)</td>
<td>bully + victim</td>
<td>journal article</td>
<td>No. Carol., USA</td>
<td>random after matching</td>
<td>small groups</td>
<td>8</td>
<td>8-9 (8.6)</td>
<td>self-report</td>
<td>381</td>
</tr>
<tr>
<td>10. Frey (2005)</td>
<td>bully + victim</td>
<td>journal article</td>
<td>Wash. state, USA</td>
<td>random after matching</td>
<td>school</td>
<td>13</td>
<td>8-12 (10.0)</td>
<td>self-report</td>
<td>1,126</td>
</tr>
<tr>
<td>11. Hunt (2007)</td>
<td>bully + victim</td>
<td>journal article</td>
<td>Sydney Australia</td>
<td>nonrandom, post hoc matching</td>
<td>school</td>
<td>0.2</td>
<td>12-15 (13.5)</td>
<td>self-report</td>
<td>400</td>
</tr>
<tr>
<td>15. Menesini (2003)</td>
<td>bully + victim</td>
<td>Journal article</td>
<td>central Italy</td>
<td>nonrandom, self-selection</td>
<td>class</td>
<td>36</td>
<td>11-14 (12.5)</td>
<td>peer report</td>
<td>293</td>
</tr>
<tr>
<td>Study (year)</td>
<td>Outcome</td>
<td>Pub. type</td>
<td>Location</td>
<td>Research design</td>
<td>Level of implementation</td>
<td>Prog. length (wks)</td>
<td>Student age in years (average age)</td>
<td>Outcome source</td>
<td>Sample size&lt;sup&gt;a,b&lt;/sup&gt; (# students)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------</td>
<td>-----------</td>
<td>----------------</td>
<td>------------------------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td>------------------------------------</td>
<td>----------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>18. Rican (1996)</td>
<td>bully + victim</td>
<td>journal article</td>
<td>Czechoslovakia</td>
<td>nonrandom, self-selection</td>
<td>class</td>
<td>12</td>
<td>9-10 (9.5)</td>
<td>self-report</td>
<td>198</td>
</tr>
</tbody>
</table>

<sup>a</sup>The studies by Baldry (2004), Rahey (2002), and Stevens (2000) involved two independent samples each, representing groups of students of different ages. Y=younger students and O=older students.

<sup>b</sup>Sample size of combined treatment and control groups, at pretest. For studies with different sample sizes for bullying and victimization outcomes, B=bullies and V=victims. Separate sample sizes are presented for the two samples in the studies by Baldry (2004), Rahey (2002), and Stevens (2000).
<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Curriculum</th>
<th>Whole school policy</th>
<th>Class rules</th>
<th>Teacher training</th>
<th>Individual work w/ bullies or victims</th>
<th>Peer mentor/ mediation</th>
<th>Parent info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alsaker (2001)</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Andreou (2007)</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Baldry (2004)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Bauer (2007)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Beran (2004)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Boulton (1996)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Cross (2004)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>DeRosier (2004a)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Fekkes (2006b)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Frey (2005)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Hunt (2007)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Jenson (2007)</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Melton (1998)</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Menard (2008)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Menesini (2003)</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Mikheev (2005)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Rahey (2002)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Rican (1996)</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Rosenbluth (2004)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Stevens (2000)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Taylor (2006)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Turpeau (1998)</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Woods (2007)</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Appendix 3
Fixed Effects Model Calculations for Bullying Outcomes Meta-Analysis

Standardized effect sizes and standard errors for each of the studies that estimated treatment effects on bullying outcomes are presented in Table 46, along with estimates of study variances and inverse variance weights under a fixed effects model. The last two columns, \( w*ES \) and \( w*ES^2 \), present calculations that are used in the derivation of the pooled effect size and the homogeneity statistic, \( Q \).

Table 46 Study-Level Data for Fixed Effects Meta-Analysis of Bullying Outcomes

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Effect size (ES)</th>
<th>Standard error (SE)</th>
<th>Variance (v)</th>
<th>Inverse variance weight (w)</th>
<th>( w*ES )</th>
<th>( w*ES^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Andreou (2007)</td>
<td>0.355</td>
<td>0.197</td>
<td>0.039</td>
<td>25.64</td>
<td>9.103</td>
<td>3.231</td>
</tr>
<tr>
<td>2 Baldry #1 (2004)</td>
<td>-0.376</td>
<td>0.297</td>
<td>0.088</td>
<td>11.36</td>
<td>-4.273</td>
<td>1.607</td>
</tr>
<tr>
<td>3 Baldry #2 (2004)</td>
<td>0.446</td>
<td>0.274</td>
<td>0.075</td>
<td>13.33</td>
<td>5.947</td>
<td>2.652</td>
</tr>
<tr>
<td>4 Boulton (1996)</td>
<td>0.075</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>1.056</td>
<td>0.079</td>
</tr>
<tr>
<td>5 Cross (2004)</td>
<td>-0.289</td>
<td>0.421</td>
<td>0.177</td>
<td>5.64</td>
<td>-1.631</td>
<td>0.471</td>
</tr>
<tr>
<td>6 DeRosier (2004a)</td>
<td>0.102</td>
<td>0.166</td>
<td>0.028</td>
<td>36.23</td>
<td>3.696</td>
<td>0.377</td>
</tr>
<tr>
<td>7 Fekkes (2006b)</td>
<td>0.073</td>
<td>0.271</td>
<td>0.073</td>
<td>13.62</td>
<td>0.994</td>
<td>0.073</td>
</tr>
<tr>
<td>8 Frey (2005)</td>
<td>0.059</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>0.831</td>
<td>0.049</td>
</tr>
<tr>
<td>9 Hunt (2007)</td>
<td>-0.110</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>-1.122</td>
<td>0.123</td>
</tr>
<tr>
<td>10 Jenson (2007)</td>
<td>0.098</td>
<td>0.142</td>
<td>0.020</td>
<td>49.75</td>
<td>4.876</td>
<td>0.478</td>
</tr>
<tr>
<td>11 Melton (1998)</td>
<td>0.217</td>
<td>0.390</td>
<td>0.152</td>
<td>6.57</td>
<td>1.427</td>
<td>0.310</td>
</tr>
<tr>
<td>12 Menard (2008)</td>
<td>0.083</td>
<td>0.272</td>
<td>0.074</td>
<td>13.51</td>
<td>1.122</td>
<td>0.093</td>
</tr>
<tr>
<td>13 Menesini (2003)</td>
<td>0.255</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>5.667</td>
<td>1.445</td>
</tr>
<tr>
<td>14 Mikheev (2005)</td>
<td>0.459</td>
<td>0.310</td>
<td>0.096</td>
<td>10.42</td>
<td>4.781</td>
<td>2.195</td>
</tr>
<tr>
<td>15 Rahey #1 (2002)</td>
<td>0.097</td>
<td>0.470</td>
<td>0.221</td>
<td>4.52</td>
<td>0.439</td>
<td>0.043</td>
</tr>
<tr>
<td>16 Rahey #2 (2002)</td>
<td>0.036</td>
<td>0.460</td>
<td>0.212</td>
<td>4.72</td>
<td>0.170</td>
<td>0.006</td>
</tr>
<tr>
<td>17 Rican (1996)</td>
<td>0.553</td>
<td>0.559</td>
<td>0.312</td>
<td>3.20</td>
<td>1.770</td>
<td>0.979</td>
</tr>
<tr>
<td>18 Rosenbluth (2004)</td>
<td>-0.006</td>
<td>0.303</td>
<td>0.092</td>
<td>10.89</td>
<td>-0.065</td>
<td>0.000</td>
</tr>
<tr>
<td>19 Stevens #1 (2000)</td>
<td>0.174</td>
<td>0.232</td>
<td>0.054</td>
<td>18.52</td>
<td>3.222</td>
<td>0.561</td>
</tr>
<tr>
<td>20 Stevens #2 (2000)</td>
<td>-0.169</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>-3.756</td>
<td>0.635</td>
</tr>
<tr>
<td>21 Taylor (2006)</td>
<td>0.088</td>
<td>0.359</td>
<td>0.129</td>
<td>7.75</td>
<td>0.682</td>
<td>0.060</td>
</tr>
<tr>
<td>22 Turpeau (1998)</td>
<td>0.096</td>
<td>0.313</td>
<td>0.098</td>
<td>10.20</td>
<td>0.980</td>
<td>0.094</td>
</tr>
<tr>
<td>( \Sigma )</td>
<td>2.316</td>
<td>6.707</td>
<td>2.271</td>
<td>328.71</td>
<td>35.914</td>
<td>15.560</td>
</tr>
</tbody>
</table>
In the fixed effects model, the pooled effect size for the prevention programs on bullying outcomes across all 22 studies is calculated as follows, using data from Table 46:

\[
\bar{ES}_{FE} = \frac{\sum_i (w_i * ES)}{\sum_i w_i} = \frac{35.914}{328.71} = 0.109.^{113}
\]

The standard error of \( \bar{ES}_{FE} \) is:

\[
SE_{\bar{ES}} = \sqrt{\frac{1}{\sum_i w_i}} = \sqrt{\frac{1}{328.71}} = 0.055
\]

Using the pooled effect size and standard error, the Z-statistic is calculated as follows:

\[
Z_{FE} = \frac{\bar{ES}}{SE_{\bar{ES}}} = \frac{0.109}{0.055} = 1.982
\]

Last, the upper and lower 95% confidence intervals (\( \bar{ES}_{FE,l} \) and \( \bar{ES}_{FE,u} \)) for the treatment effect are constructed using the pooled effect size and the critical Z-value representing the desired confidence level (in this case, \( \alpha=.05 \) and \( Z=1.96 \)):

\[
\bar{ES}_{FE,l} = \bar{ES}_{FE} - 1.96(SE_{FE}) = 0.109 - 1.96(0.055) = 0.001
\]
\[
\bar{ES}_{FE,u} = \bar{ES}_{FE} + 1.96(SE_{FE}) = 0.109 + 1.96(0.055) = 0.217
\]

\(^{113}\) Due to rounding error, the coefficients have slight differences from those presented in the body of the paper.
Since the confidence interval does not include zero (and the Z-statistic of 1.98 exceeds the critical Z-value of 1.96), we reject the null hypothesis of no effect. The pooled effect for bullying prevention programs on bullying outcomes is statistically significant.
Appendix 4
Q-Statistic for Bullying Outcomes Meta-Analysis

To calculate the $Q$-statistic for homogeneity, we refer to the values $\sum w_i$, $\sum w_i^* ES_i$, and $\sum w_i^* ES_i^2$ from Table 46 (Appendix 3). The $Q$-statistic for the 22 effect sizes measuring bullying outcomes is calculated as:

$$Q = \frac{\sum_i (w_i^* ES_i^2) - \frac{\sum_i (w_i^* ES_i)^2}{\sum_i w_i}}{\sum_i w_i} = 15.560 - \frac{(35.914)^2}{328.71} = 15.560 - \frac{1289.82}{328.71} = 11.636.$$

The $Q$-value of 11.6 on 21 degrees of freedom ($p = 0.950$) suggests that the studies are estimating the same population parameter.

The $I^2$ statistic is used to assess the magnitude of heterogeneity:

$$I^2 = \frac{Q - (k - 1) \times 100}{Q} = \frac{11.6 - (22 - 1)}{11.6} = -0.810 = 0\% \text{ (negative values of } I^2 \text{ are set to 0).}$$

An $I^2$ of 0\% indicates no meaningful heterogeneity among studies in the included set.

---

114 This formula is algebraically equivalent to the formula for $Q$ presented in the Analysis section, and is simpler to implement for the purposes of hand calculations.

115 Higgins et al., 2003.
Appendix 5
"Research Design" Subgroup Analysis for Bullying Outcomes Meta-Analysis

Following is the step-by-step subgroup analysis procedure for the variable "research design", as calculated for the bullying meta-analysis. The coefficients differ slightly from those presented in the body of the report (see Table 33); these differences are due to rounding error.

(1) First, the previously defined variables (effect size ($ES$), standard error ($SE$), variance ($v$), inverse variance weight ($w$), $w*ES$ and $w*ES^2$) are summed separately for each category of the predictor variable (randomized versus non-randomized research design), as shown in Table 47:
Table 47 Study-Level Data by Research Design Subgroup for Bullying Outcomes

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>Effect size (ES)</th>
<th>Standard error (SE)</th>
<th>Variance (v)</th>
<th>Inverse variance weight (w)</th>
<th>w*ES</th>
<th>w*ES^2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Randomized studies</strong> (R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Baldry #1 (2004)</td>
<td>-0.376</td>
<td>0.297</td>
<td>0.088</td>
<td>11.36</td>
<td>-4.273</td>
<td>1.607</td>
</tr>
<tr>
<td>2 Baldry #2 (2004)</td>
<td>0.446</td>
<td>0.274</td>
<td>0.075</td>
<td>13.33</td>
<td>5.947</td>
<td>2.652</td>
</tr>
<tr>
<td>3 Boulton (1996)</td>
<td>0.075</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>1.056</td>
<td>0.079</td>
</tr>
<tr>
<td>4 Cross (2004)</td>
<td>-0.289</td>
<td>0.421</td>
<td>0.177</td>
<td>5.64</td>
<td>-1.631</td>
<td>0.471</td>
</tr>
<tr>
<td>5 DeRosier (2004a)</td>
<td>0.102</td>
<td>0.166</td>
<td>0.028</td>
<td>36.23</td>
<td>3.696</td>
<td>0.377</td>
</tr>
<tr>
<td>6 Fekkes (2006b)</td>
<td>0.073</td>
<td>0.271</td>
<td>0.073</td>
<td>13.62</td>
<td>0.994</td>
<td>0.073</td>
</tr>
<tr>
<td>7 Frey (2005)</td>
<td>0.059</td>
<td>0.266</td>
<td>0.071</td>
<td>14.08</td>
<td>0.831</td>
<td>0.049</td>
</tr>
<tr>
<td>8 Jenson (2007)</td>
<td>0.098</td>
<td>0.142</td>
<td>0.020</td>
<td>49.75</td>
<td>4.876</td>
<td>0.478</td>
</tr>
<tr>
<td>9 Rosenbluth (2004)</td>
<td>-0.006</td>
<td>0.303</td>
<td>0.092</td>
<td>10.89</td>
<td>-0.065</td>
<td>0.000</td>
</tr>
<tr>
<td>10 Stevens #1 (2000)</td>
<td>0.174</td>
<td>0.232</td>
<td>0.054</td>
<td>18.52</td>
<td>3.222</td>
<td>0.561</td>
</tr>
<tr>
<td>11 Stevens #2 (2000)</td>
<td>-0.169</td>
<td>0.212</td>
<td>0.045</td>
<td>22.22</td>
<td>-3.756</td>
<td>0.635</td>
</tr>
<tr>
<td>12 Taylor (2006)</td>
<td>0.088</td>
<td>0.359</td>
<td>0.129</td>
<td>7.75</td>
<td>0.682</td>
<td>0.060</td>
</tr>
<tr>
<td><strong>Σ (n=12)</strong></td>
<td><strong>0.275</strong></td>
<td><strong>3.210</strong></td>
<td><strong>0.923</strong></td>
<td><strong>217.49</strong></td>
<td><strong>11.579</strong></td>
<td><strong>7.041</strong></td>
</tr>
</tbody>
</table>

| Non-randomized studies (NR) |                  |                     |              |                            |        |        |
| 1 Andreou (2007)          | 0.355            | 0.197               | 0.039        | 25.64                      | 9.103  | 3.231  |
| 2 Hunt (2007)             | -0.110           | 0.313               | 0.098        | 10.20                      | -1.122 | 0.123  |
| 3 Melton (1998)           | 0.217            | 0.390               | 0.152        | 6.57                       | 1.427  | 0.310  |
| 4 Menard (2008)           | 0.083            | 0.272               | 0.074        | 13.51                      | 1.122  | 0.093  |
| 5 Menesini (2003)         | 0.255            | 0.212               | 0.045        | 22.22                      | 5.667  | 1.445  |
| 6 Mikheev (2005)          | 0.459            | 0.310               | 0.096        | 10.42                      | 4.781  | 2.195  |
| 7 Rahey #1 (2002)         | 0.097            | 0.470               | 0.221        | 4.52                       | 0.439  | 0.043  |
| 8 Rahey #2 (2002)         | 0.036            | 0.460               | 0.212        | 4.72                       | 0.170  | 0.006  |
| 9 Rican (1996)            | 0.553            | 0.559               | 0.312        | 3.20                       | 1.770  | 0.979  |
| 10 Turpeau (1998)         | 0.096            | 0.313               | 0.098        | 10.20                      | 0.980  | 0.094  |
| **Σ (n=10)**              | **2.041**        | **3.497**           | **1.348**    | **111.22**                 | **24.334** | **8.519** |
| **TOTAL Σ (n=22)**        | **2.316**        | **6.707**           | **2.271**    | **328.711**                | **35.914** | **15.560** |
(2) Second, the pooled effect size and standard error are computed for each group (randomized and non-randomized) in the same manner as they were for the full set of effect sizes (see Appendix 3).

(a) Randomized studies:

\[
\bar{ES}_R = \frac{\sum_i (w_i \cdot ES)}{\sum_i w_i} = \frac{11.579}{217.49} = 0.053
\]

\[
SE_{ES_R} = \sqrt{\frac{1}{\sum_i w_i}} = \sqrt{\frac{1}{217.49}} = 0.068
\]

(b) Non-randomized studies:

\[
\bar{ES}_{NR} = \frac{\sum_i (w_i \cdot ES)}{\sum_i w_i} = \frac{24.334}{111.22} = 0.219
\]

\[
SE_{ES_{NR}} = \sqrt{\frac{1}{\sum_i w_i}} = \sqrt{\frac{1}{111.22}} = 0.095
\]
(3) The $Q$-statistic for homogeneity is then calculated for the randomized study subgroup (12 studies) and the non-randomized subgroup (10 studies):

(a) $Q_r = \left( \sum_i w_i ESI_i^2 \right) - \frac{\sum_i (w_i * ESI_i)^2}{\sum_i w_i} = 7.041 - \frac{(11.579)^2}{217.49} = 7.041 - \frac{134.07}{217.49} = 6.425$

The $Q$-statistic for the randomized studies = 6.425, distributed as a chi-square with 11 degrees of freedom (p=0.844).

(b) $Q_{nr} = \left( \sum_i w_i ESI_i^2 \right) - \frac{\sum_i (w_i * ESI_i)^2}{\sum_i w_i} = 8.519 - \frac{(24.334)^2}{111.22} = 8.519 - \frac{592.144}{111.22} = 3.195$

The $Q$-statistic for the non-randomized studies = 3.195, distributed as a chi-square with 9 degrees of freedom (p=0.956).

(4) Next, the within-group and between-groups homogeneity statistics are calculated.

(a) The pooled within-groups homogeneity statistic, $Q$-within ($Q_{\text{w}}$), is defined as the sum of the subgroup $Q$s:

$Q_{\text{w}} = Q_r + Q_{\text{nr}} = 6.425 + 3.195 = 9.62.$

Degrees of freedom = $k - j$, where $k$ is the number of effect sizes and $j$ is the number of groups, therefore, $df = (12+10) - 2 = 20$. 
(b) The total $Q^{116}$ is partitioned in two, a within-group $Q$ and a between-groups $Q$. The difference between $Q$-total ($Q_t$) and $Q$-within ($Q_w$) is $Q$-between ($Q_b$):

$$Q_b = Q_t - Q_w = 11.636 - 9.62 = 2.016.$$  

The degrees of freedom for $Q_b$ is $j - 1$, therefore, $df = 1$.

(5) The $Q_B$ and $Q_W$ estimates are then tested. For $Q_B = 2.016$, $df = 1$ and $p = .156$. The critical value of chi-square at $\alpha = .05$ and $df = 1$ is 3.84. Since $Q_B < Q_{critical}$, the mean effect sizes produced by each subgroup are not considered to be statistically significantly different from one another.

For $Q_W = 9.62$, $df = 20$ and $p = .975$, the critical value of chi-square at $\alpha = .05$ and $df = 20$ is 31.410. Since $Q_W < Q_{critical}$, $Q_W$ is not statistically significant, this finding suggests that, after accounting for the between-groups heterogeneity, the residual variability within each subgroup is not greater than we would expect by chance.

---

116 $Q$-total is the $Q$-statistic across all studies, defined in Appendix 4.