PREVENTING AGGRESSIVE BEHAVIOR BY PROMOTING SOCIAL INFORMATION-PROCESSING SKILLS: A THEORY-BASED EVALUATION OF THE *MAKING CHOICES* PROGRAM

Mary Arminé Terzian

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in the School of Social Work.

Chapel Hill
2007

Approved by:

Mark Fraser

Paul Smokowski

Shenyang Guo

Natasha Bowen

Shari Miller-Johnson

ABSTRACT

MARY TERZIAN: Preventing Aggressive Behavior by Promoting SIP Skills:
A Theory-based Evaluation of the Making Choices Program
[Under the direction of Mark Fraser, Chair, and Natasha Bowen, Advisor]

This theory-based evaluation was conducted on pretest-posttest data collected from an efficacy trial of the *Making Choices (MC)* program, a universal intervention designed to prevent conduct problems. This study examined three areas of inquiry. First, program effects on social information-processing (SIP) skills and overt aggression were evaluated. Next, program-by-gender interactions were tested. Lastly, indirect effects were tested to evaluate whether effects on theoretical mediators, in part, explained program success. *MC* and *MC+* were expected to result in decreased overt aggression and improved SIP skills, and gender was expected to moderate these effects. SIP skills were expected to partially mediate program effects on overt aggression.

The study utilized a non-randomized, cohort design with treatment withdrawal. The sample consists of three ethnically-diverse cohorts of third graders (N=480; 50% female) from two rural elementary schools. The 2001-02 cohort (n=156) participated in *MC,* the 2002-03 cohort (n=193) participated in *MC+,* an augmented version of *MC*. After a one-year treatment withdrawal period, data were collected from a routine-services cohort (2004-05; n=131).

On average, intervention students made greater improvements than comparison students on all outcomes. *MC* students demonstrated better encoding, emotion regulation, and

response selection. *MC*+ students had less hostile attribution bias and better emotion regulation and response selection. *MC* and *MC*+ boys had less overt aggression and more benign social goals than comparison boys. *MC*+ girls also experienced improvements on these outcomes, though improvements were modest. Effect sizes for SIP skills varied in magnitude from small to medium, and large effects on overt aggression were obtained for boys. Three out of five SIP skills (i.e., goal clarification, response selection, and emotion regulation) appeared to explain program effects on overt aggression.

Effects on social cognition were consistent with study hypotheses. Large effects on overt aggression suggest that SIP-based programs may be particularly effective at preventing conduct problems in boys. Mediation findings suggest that effects on SIP skills explained program effects. Intervention research examining mediation and moderation can help us to achieve a better understanding of "what works" and "for whom," improving our capacity to prevent problem behavior in youth.

DEDICATION

To my parents, Makrouhi and Carnig, and to my sister Arpi.

Your encouragement and support have helped make this achievement possible.

To my faculty mentors: Professors Fraser, Galinsky, Smokowski, Bowen, and Guo.

And to all of my friends at the UNC-CH School of Social Work.

You have made my time in Chapel Hill memorable and worthwhile.

ACKNOWLEDGEMENTS

I am fortunate to have had the opportunity to work with many talented scholars on my dissertation research. First, I am indebted to Dr. Mark Fraser, the Principal Investigator of the *Making Choices* project and the chair of my dissertation committee. His conceptual and methodological guidance helps me to conduct rigorous research that answers meaningful questions and has practical applications. Second, I would like to thank my committee members: Paul Smokowski, Shenyang Guo, Natasha Bowen, and Shari Miller-Johnson, for their thoughtful comments and critiques. Third, I must thank Steve Day, Evaluation Director of the Carolina Children's Initiative, for orienting me to the project and familiarizing me with data collection/scoring protocols, and Kathlene Holmes, for assisting with data collection. Fourth, I would like to acknowledge the assistance I received from statisticians Roderick Rose, Pat Malone, and Chongming Yang, from UNC-Chapel Hill and Duke University. Finally, I am indebted to the Doctoral Program Chair, Dr. Kathleen Rounds, and to the former Chair, Dr. Richard Barth, for supporting my dissertation research and for encouraging my academic and professional development.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AA	African American
α	Alpha
ACP	*Anger Coping Program*
ANOVA	Analysis of Variance
β	Standardized beta coefficient
B	Unstandardized beta coefficient
CBCL	*Child Behavior Checklist*
CC	Comparison Condition
CCC-TF	*Carolina Child Checklist – Teacher Form*
χ^2	Chi-square
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CPPRG	Conduct Problems Prevention Research Group
δ	Standardized Delta
df	Degrees of Freedom
DHHS	Department of Health and Human Services
EBD	Emotional and Behavioral Disturbance
ES	Effect Size
ESL	English as a Second Language
GLM	Generalized Linear Model
HLM	Hierarchical Linear Model

IEP	Individualized Education Program
IOM	Institute of Medicine
κ	Kappa
LSD	Least Significant Difference
MANOVA	Multiple Analysis of Variance
MC	*Making Choices*
MC+	*Making Choices Plus*
MC+SF	*Making Choices plus Strong Families*
ML	Maximum Likelihood
MSEM	Multilevel Structural Equation Modeling
NICHD	National Institute of Child Health and Human Development
ω	Omega
OLS	Ordinary Least Squares
π	Pi
PATHS	*Promoting Alternate Thinking Strategies*
REML	Restricted Maximum Likelihood
ρ	Rho
RMSEA	Root Mean Square Error of Approximation
SAS	Statistical Analysis System
sbp	Spearman Brown Prophecy
SHP	*Social Health Profile*
σ	Sigma
SIP	Social Information Processing

SLA	*Skill Level Activity*
SEM	Structural Equation Modeling
SES	Socioeconomic status
SPS	Social Problem Solving
√	Square Root
t	*t* statistic
τ	Tau
T1	Time 1 (pretest)
T2	Time 2 (posttest)
TLI	Tucker-Lewis Index
TOCA-R	*Teacher Observation of Classroom Adaptation - Revised*
TRF	*Teacher Report Form*
U.S.	United States
YMCA	Young Men's Christian Association
Z	*Z* statistic

Chapter I

Introduction

Among the constellation of individual-, familial-, and peer-level risk factors implicated in the development of problem behavior in youth, childhood overt aggression may be the most commonly identified (Coie, Lochman, Terry, & Hyman, 1992; Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998). Overt aggression relates to confrontational verbal and/or physical behavior that adversely affects others and often leads to rejection by same-age peers and an increased vulnerability to peer victimization (Hanish & Guerra, 2000; Kochenderfer-Ladd, 2003). Aggressive-rejected children appear to have a higher risk of poor developmental outcomes than aggressive-nonrejected children (Bierman & Wargo, 1995; Miller-Johnson, Coie, Maumary-Gremaud, & Bierman, 2002; Prinstein & LaGreca, 2004). Nonetheless, regardless of a child's rejection-status, overt aggression in childhood appears to uniquely predict later conduct problems, academic failure, teenage pregnancy, drug abuse, and juvenile delinquency (Coie et al., 1992; Coie, Malone, & Lochman, 2004; Nagin & Tremblay, 1999; Schwartz, McFayden-Ketchum, Dodge, Pettit, & Bates, 1999).

Aggression in early childhood is particularly predictive of antisocial behavior. Youths who began to display aggressive behavior as toddlers and preschoolers are commonly termed *early-start* youths (Patterson & Yoerger, 1993). Early-start youths often reside in families where parent-child exchanges are coercive, demonstrate poor school adjustment, and, by approximately the third grade, begin to experience rejection from their prosocial peers (Pope,

Bierman, & Mumma, 1989). They are also at heightened risk for embarking on a *life-course-persistent* trajectory of criminal offending (Moffit, 2003; Patterson, Forgatch, Yoerger, & Stoolmiller, 1998). These youths represent about 6% of the general population, but they perpetrate nearly half of all adolescent crimes (Conduct Problems Prevention Research Group [CPPRG], 1999a).

In the last few decades, much progress has been made toward the understanding, prevention, and treatment of aggressive behavior in children and adolescents. Research in developmental psychopathology and psychiatric epidemiology has suggested multiple points of intervention for youth of different ages, risk levels, and symptom profiles (Kellam & Rebok, 1992; Lochman & Wells, 2002; Miller-Johnson et al., 2002; Moffit & Caspi, 2001; Vitaro, Brendgen, Pagani, Tremblay, & McDuff, 1999).

Rooted in a risk and resilience perspective, which conceives human development as on ongoing interaction between personal and environmental resources and deficits (Rutter, 1997; Sameroff, 1985), multi-element violence prevention programs attempt to reduce aggression by targeting malleable aspects of the person (e.g., cognitive, social-emotional, and academic skills) and environment (e.g., classroom management, peer norms, school climate, parental discipline and monitoring) (CPPRG, 1999a; O'Donnell, Hawkins, Catalano, Abbott, & Day; 1995; Reid, Eddy, Fetrow, & Stoolmiller, 1999). Encouraged by the success of classroom-based skills training programs such as *Second Step* (Grossman et al., 1997), *I Can Problem Solve* (ICPS; Spivack & Shure, 1985), and *Promoting Alternative Thinking Strategies* (*PATHS*; Greenberg, Kusché, Cook, & Quamma, 1995), multi-element school-based programs commonly seek to improve social-cognitive skills and emotional competence (Bierman et al., 2002; Kazdin, Siegel, & Bass, 1992; Losel & Beelman, 2003; Reid et al.,

1999). By and large, these programs have been effective at deterring aggression in elementary school children and several have led to long-term social and behavioral improvements (Losel & Beelman, 2003; Terzian & Fraser, 2005).

Indeed, youth development programs such as these have contributed to our knowledge about what is effective for normative, at-risk, and indicated samples of youth. First, we have learned that interventions conducted with homogeneous groups of aggressive children are likely increase problem behavior (Poulin, Dishion, & Burraston, 2001; Tolan, 2001), unless combined with a multi-element approach that engages children with prosocial peers in the classroom context (CPPRG, 1999a). Second, we have learned that involving families in school-based programs is a challenging but worthwhile task (Farrington & Welsh, 1999; Tremblay, Pagani-Kurtz, Mâsse, Vitaro, & Pihl, 1995). Third, we have learned that implementation issues such as service intensity, duration, dosage, and treatment fidelity make a difference and must be monitored and evaluated for programs to be effective (Abbot, O'Donnell, Hawkins, Hill, Kosterman & Catalano, 1998; Mihalic, Irwin, Fagan, Ballard, & Elliot, 2004). And fourth, we have learned that organizational and system-level factors are vital to the implementation, evaluation, and sustainability of the program (Greenberg et al., 2003; Ozer, 2006).

But there is still more to learn. To date, few intervention research studies have conducted theory-based, process-oriented evaluations to decompose explanatory mechanisms underlying program effects. Instead, most have focused solely on testing main effects of study outcomes. Without testing the impact of theoretical mediators, mechanisms leading to behavioral improvement remain hidden inside the 'black box' (Kazdin & Nock, 2003; Lipsey, 1988). The lack of precision in evaluation could have serious implications for

community-based intervention research. School-based intervention research can be expensive and difficult to implement. Lacking knowledge of program mechanisms, schools attempting to implement 'effective' interventions in good faith may unintentionally alter critical ingredients necessary for program success. As a result, programs that have been deemed 'evidence-based' may yield null effects when brought to scale. Theory-based evaluations that test the effects of theoretical mediators can facilitate the identification of mechanisms responsible for social and behavioral improvements (Birckmayer & Weiss, 2000; Howe, et al., 2002).

This dissertation is a theory-based evaluation of the *Making Choices* (*MC*) program, a universal preventive intervention designed to reduce aggression in elementary school-age children (Fraser, Nash, Galinsky, & Darwin, 2000). A recently published study (Fraser et al., 2005) found that the *MC* program had effects on posttest overt aggression. The current study seeks to re-evaluate main effects on posttest outcomes, determine whether gender moderates these effects, and determine whether theory-based constructs account for program effects on overt aggression. Five social information-processing (SIP) variables were tested as potential mediators: a) encoding; b) hostile attribution (interpretation); c) goal clarification; d) response selection (i.e., a component of response decision); and e) emotion regulation. These mediators were hypothesized to partially explain program effects on posttest ratings of overt aggression.

In summary, this dissertation sought to answer the following research questions:

- Does *MC* and *MC*+ have effects on overt aggression and SIP Skill?
- Do direct and indirect effects vary by gender? and
- Does SIP skill explain program effects on overt aggression at posttest?

Answering these questions may yield relevant and valuable information for social work practice. For example, testing indirect program effects via social-cognitive skills may help to validate the promotion of social-cognitive skills as an intervention strategy for reducing aggression in elementary-school children. Also, these mediational tests can offer information about whether certain social-cognitive skills have more relevance than others at this age. Testing the moderating effects of gender allows researchers to estimate differential effects of the program on boys' and girls' social cognition and behavior and discern the relative importance of different SIP skills and processes in the enactment of aggression. Finally, studies that more fully examine the moderating role of gender on aggressogenic (i.e., aggression-causing) processes can assist social work practitioners in the design and development of gender-specific delinquency prevention programs.

Statement of the Problem

Youth violence is a serious problem impacting schools and communities throughout the United States. Public concern related to the problem of youth violence mounted in the early-1990s, when rates of violent juvenile crime reached record highs (Office of Juvenile Justice and Delinquency Prevention, 2002). Alarmingly, the mean age of the juvenile offender population dropped, and arrest rates for violent crime increased more rapidly for females than for males (McCabe, 2002). The female population of juvenile offenders was especially young, with over 30% being under the age of 15 (Chesney-Lind, 1999). Federal agencies, including the U.S. Department of Justice and the U.S. Department of Health and Human Services, responded to these disturbing trends by sponsoring large-scale intervention studies, launching gender-specific programming initiatives, and funding epidemiological

studies of clinical and non-referred samples. To be sure, preventing youth violence in the U.S. has become a national priority.

This chapter will examine the epidemiology of childhood aggression. First, the prevalence of overt aggression in the U.S. is described, according to subtype, age, gender, and risk-status. Second, developmental outcomes associated with overt aggression in early and middle childhood are considered. Finally, the chapter will discuss the significance of overt aggression for social work practice. In order to fully understand the problem of overt aggression, the definition of this problem must first be clarified.

Defining Aggression

The study of childhood aggression has led to the conceptualization of aggression as a multifaceted construct that is composed of multiple, yet overlapping, subtypes, each posited to relate to a distinct etiology and developmental trajectory (Dubow, Huesmann, & Boxer, 2003; Vitaro, Gendreau, Tremblay, & Oligny, 1998). Subtypes of aggression have been characterized by function (instrumental-relational), intention (reactive-proactive), and behavior type (direct-indirect; overt-covert; physical-verbal-social) (Dodge & Coie, 1987; Little, Jones, & Henrich., 2003). Common to all forms of aggression is behavior that inflicts aggravation, injury, and/or harm to others.

Overt aggression involves the use of direct verbal and/or physical confrontation. It occurs either in response to provocation from others (*reactive*) or is strategically directed toward others (*proactive*). Proactive overt aggression includes but is not subsumed by behaviors commonly associated with bullying, such as name-calling, teasing, threatening, and shoving. Unlike bullying, however, overt aggression may not be directed toward a victim who is perceived to have less social and/or physical power (Limber & Nation, 1998).

Reactive overt aggression is characterized by behaviors such as displaying outbursts of anger, overreacting easily to situations, and blaming others in fights. Dodge and Coie (1987) constructed a scale measuring reactive and proactive aggression. Items on this scale loaded onto two factors which were strongly correlated with each other (r=.76). These factors had convergent validity with direct observational measures and their reported internal consistencies were high (i.e., r =.91 for proactive aggression and r =.90 for reactive aggression). More recent studies have replicated the finding that proactive and reactive aggression comprises two distinct but highly related constructs (r=.82; Poulin & Boivin, 2000). The strength of the correlation between these two subtypes may vary depending the function of the behavior (e.g., overt or indirect; Prinstein & Cilessen, 2003).

Overt aggression in children is typically measured using teacher-rated and parent-rated questionnaires, often as a subscale within a broader measure of problem behavior, such as the externalizing scale of the *Child Behavior Checklist* (*CBCL*; Achenbach & Edelbrock, 1991), or within a measure of disruptive behavior that includes hyperactive, impulsive behavior, such as the *School Social Behavior Scales* (*SSBS*; Merrell, Sanders, & Popinga, 1993) and the *Breyer's Behavior Observation Schedule for Pupils and Teachers* (*BOSPT*; Breyer & Calchera, 1971). The *Modified Overt Aggression Scale* (*MOAS*; Kay et al., 1988) assesses the nature and prevalence of aggression in the psychiatric population. Overt aggression is also measured using peer nominations, self reports, and behavioral observations. Self-report methods range from questionnaire, structured and open-ended interviews, and responses to hypothetical vignettes involving peer provocation. For school-age children, teacher ratings may be a more reliable method of measuring aggression than

child report. Peer nominations of aggression are also known to be highly reliable (Asher & Hymel, 1981). Ideally, multiple methods and informants for measuring aggression are used.

The Prevalence of Childhood Overt Aggression

Few epidemiological studies of overt aggression in U.S. youth exist, however estimates obtained from a collection of studies indicate that prevalence rates of overt aggression are high (Nansel, Overpeck, Pilla, Ruan, Simons-Morton, & Scheidt, 2001). In the largest epidemiological survey of bullying conducted thus far in the U.S. (conducted with 15,868 sixth-through-tenth grade youth), approximately 30% of youth reported having involvement in bullying (Nansel et al., 2001). Clinical forms of aggressive, antisocial behavior in youth tend to have lower estimates. Prevalence rates for disruptive behavior disorders in U.S. youth (ages 6 to 18; N=1641) were estimated using parent-rated data collected over three decades (Achenbach, Dumenci, & Rescorla, 2003). In 1999, about 9% of the study sample reached clinical thresholds for oppositional behavior and 7% reached clinical thresholds of conduct problems, on the *CBCL* (Achenbach et al., 2003).

Gender differences in prevalence rates for aggressive and antisocial behavior exist in normative samples, but narrow significantly in clinical samples of youth (Connor, Steingard, Anderson, & Melloni, 2003). Age and gender differences in aggression were reported in a recent household survey of 1,285 youths aged 9- to 17-years-old (Lahey et al., 2000). Graphs depicting levels of aggression by age show a quadratic trend with children in the middle age range (ages 11 to14) having higher levels of aggression than children in the younger (ages 9 to 10) and older age ranges (ages 15 to 17). Although these trends differed slightly by gender (boys had a bimodal trend and girls had a unimodal trend), the quadratic term for age was significant ($p<.05$) after controlling for gender and the linear term for age. *Figure 1* displays

the percentage of boys and girls who, according to combined parent and youth reports, engaged in at least one aggressive behavior in the past 12 months.

Figure 1: Prevalence of Aggressive Behavior by Age and Gender

These findings suggest that, without intervention, the prevalence of aggression in boys is likely to escalate until age 14 and the prevalence of aggression in girls is likely to escalate until age 13. Intervening before this point may lower risk of maladjustment in adolescence.

Childhood Aggression as a Precursor to Later Maladjustment

Childhood aggression is a developmental risk factor that appears to exert crosscutting effects on girls and boys from diverse racial and cultural backgrounds (Coie et al., 1992; Côté, Zoccolillo, Tremblay, Nagin, & Vitaro, 2001; Moffit & Caspi, 20001; Prinstein & LaGrecam, 2004; Shaeffer, Petras, Ialongo, Poduska, & Kellam, 2003). Moffit and Caspi (2001) examined gender differences in the developmental trajectories of a predominantly Caucasian sample of children in Great Britain (N=1,037) and found that overt aggression at

age 3 characterized the risk profile of the life course persistent (LCP) trajectory for both genders. This finding was replicated in a study conducted with an ethnically diverse sample of 146 elementary school youth located in a large urban metropolitan area in the U.S. (Prinstein & LaGreca, 2004). Though peer acceptance moderated the predictive power of teacher-rated overt aggression, girls and boys who displayed overt aggression in the fourth-, fifth-, or -sixth grades were at greater risk of externalizing problems six years later. A longitudinal study by Côté et al. (2001) in Quebec grouped 820 females (ages 14 to 17; 97% White) by their level of childhood disruptive behavior (i.e., low, medium, medium-high, and high) and tracked these four groups over ten years. Gender-specific cut-offs were used to categorize levels of aggression; girls rated as having high aggression in kindergarten had to score above the 80th percentile on 13 items from the Social Behavior Questionnaire (SBQ; Tremblay et al., 1991), a measure of antisocial, aggressive, and disruptive/hyperactive behavior. Girls in the medium-high and high groups were 4.5 times more likely than girls in the low group to have a diagnosis of Conduct Disorder (CD) and 64.3% (18 out of 28) of girls with a CD diagnosis in adolescence had medium-high to high levels of disruptive behavior during elementary school.

Childhood aggression has also been linked to later antisocial behavior in samples of African American youth (Coie et al., 1992; Shaeffer et al., 2003). Shaeffer et al. (2003), for instance, conducted a 7-year longitudinal study with an urban, community-based sample of predominantly African American boys (n=205) who were first assessed at age six as part of a study evaluating a universal, school-based intervention. The study found that boys in the chronic high and increasing aggression trajectories had higher rates of juvenile and adult arrest, conduct disorder, and antisocial personality disorder.

Indeed, overt aggression in early and middle childhood is a major risk factor that appears to negatively impact the developmental outcomes of girls and boys from diverse racial, ethnic, and cultural backgrounds. Moreover, its prevalence rate among normative samples of youth is relatively high. In light of its prevalence and widespread effects, overt aggression is an excellent target for primary prevention programs.

Significance of Problem to Social Work Practice

Childhood aggression is associated with a range of problems that often require the coordinated response of service systems – child welfare, education, juvenile justice, mental health, public health, social services, and substance abuse. As we have seen, bullying and chronic peer victimization can lead to dangerous school climates that can spark violence and even traumatic events such as school shootings and suicides that affect everyone involved. As a result of heightened public awareness and a corresponding wave of state and federal initiatives to tackle the problem of youth violence, many school systems are now beginning to implement anti-bullying campaigns that support a zero-tolerance policy for aggression (Department of Health and Human Services, 2004). These campaigns often utilize principles of character education (e.g., respecting others and being a good friend) and strategies for social development (e.g., calming down and evaluating potential responses to social problems) (Pelham, Massetti, & Waschbusch, 2005). School social workers are beginning to play an important role in delinquency prevention (Lindsay & Kurtz, 1987). Equipped with knowledge and skills about how to effectively intervene with aggressive youth, these social work professionals have the potential to play an important role in this nation-wide response.

In addition, social workers are often involved in the assessment and treatment of aggressive, antisocial behavior. Increasingly, schools are becoming a primary setting for the

delivery of mental health services to youth (Burns, et al., 1995). Individualized Education Programs (IEPs) for youth with emotional and behavioral disturbance (EBD) commonly involve school social workers and community-based mental health agencies, and aggressive children make up a significant proportion of these youth. Social work service professionals may be involved in the development of the IEP or be asked to join the IEP team as a person "with knowledge or special expertise about the child" (Office of Special Education and Rehabilitative Services, 2000).

Frequently, aggressive youth who do not respond to school- and community-based interventions, and youth who lack needed services, become involved in delinquent and/or criminal behavior, and after repeated violations, become adjudicated as wards of the state. Youth under the age of 18 are responsible for 15% of all arrests for violent crime (Federal Bureau of Investigation, 2002) and youth with conduct problems comprise a significant portion of the juvenile justice population (Vermeiren, Jespers, & Moffitt, 2006). The six-month prevalence of Conduct Disorder was estimated at 32%, in one sample of 320 Illinois and New Jersey youth (Wasserman, Ko, & McReynolds, 2004). Another report, conducted on detained Virginia youth, estimated prevalence rates of up to 52% for Disruptive Behavior Disorders (Policy Design Team, 1994).

Finally, approximately 580,000 youth are involved in the child welfare system (Leslie, Gordon, Lambros, Premji, Peoples, & Gist, 2005). A significant proportion of these youth exhibit problem behaviors (Wall, 2005). A National Study of Child and Adolescent Wellbeing (NSCAW) survey conducted with a sample of 3,803 youth involved in the child welfare system (ages 2 to 14) found 48% of this sample to have clinically significant emotional and behavioral problems (Burns et al., 2004). To meet the mental health needs of

these youth and reduce stress on biological and foster families, child welfare providers must refer children to in-home and outpatient social work services.

Without a doubt, the development and design of preventive interventions to reduce childhood aggression is a critical task for social work research and practice. To effectively prevent and treat antisocial behavior, social workers must have an adequate understanding of the epidemiology and etiology aggression. Theories of aggression have emerged from numerous fields (e.g., social work, psychology, sociology, and criminology). The following chapter reviews a specific social-cognitive framework of aggression that has informed a number of interventions to prevent aggressive behavior in youth.

Chapter II

Aggression from a Social-Cognitive Perspective

Research examining the correlates and precursors of aggression has implicated the role of malleable, social-cognitive and emotional structures and processes, such attitudes and beliefs, problem solving skill, emotional understanding, and emotion regulation in the initiation of aggression (Crick and Dodge, 1994; Hubbard, 2001; Huesmann, 1998). Social-cognitive frameworks for aggressive behavior have tended to concentrate on particular aspects of social cognition, by highlighting the role of knowledge structures (Burks, Laird, Dodge, Pettit, & Bates., 1999; Huesmann, 1988), attribution (Graham, Hudley, & Williams, 1992), emotion-related processes (Eisenberg et al., 2001), or problem solving skills (D'Zurilla & Goldfried, 1971). Recently, scholars have begun to develop more unified theories that integrate multiple aspects of social cognition (Crick & Dodge, 1994; Huesmann, 1998; Lemerise & Arsenio, 2000).

One such theoretical framework is the social-information processing (SIP) model (Dodge, 1986; Crick & Dodge, 2004). This model has been applied to other problem conditions, such as depression (Quiggle, Garber, Panak, & Dodge, 1992) and attention-deficit and hyperactivity disorder (Cadesky, Mota & Schachar, 2000). Empirical support for this model as an explanation for aggression has accrued from numerous studies, ranging from experimental studies in developmental psychopathology (Orobio de Castro, Merk, Koops, Veerman, & Bosch, 2005) to school- and community-based intervention research (Fraser et

al., 2005; Grossman et al., 1997; Hudley & Graham, 1993; Lochman & Wells, 2002). The following section discusses the reformulated SIP model and reviews its empirical support.

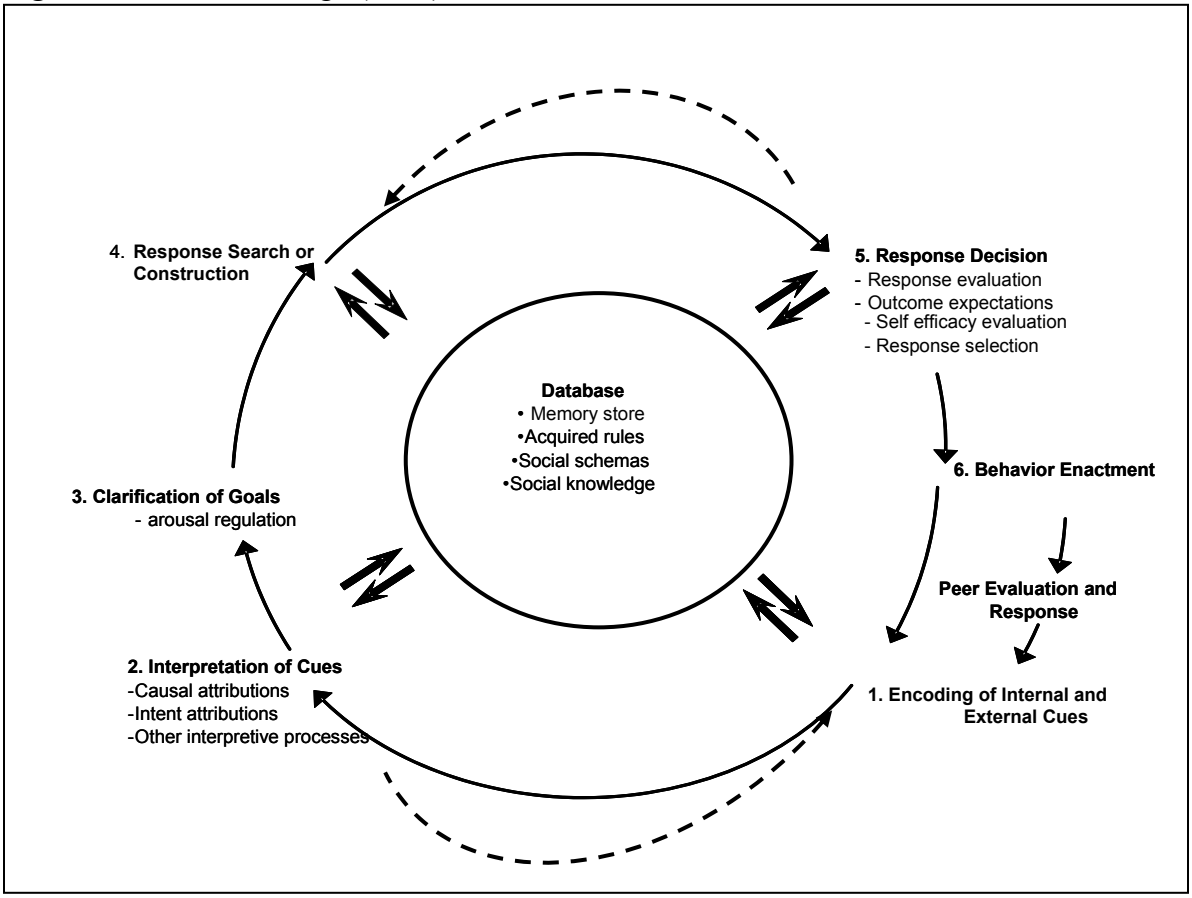
The Reformulated Social Information-Processing Model (SIP)

The *Making Choices* program is guided by an emotion-integrated, social cognitive perspective, which is based on a body of research that links peer rejection and aggression to the way in which children perceive, interpret, evaluate, and respond to social situations (Fraser, 1996). This conceptualization of aggression is called the *social information-processing model*. Central to the SIP model is the idea that one's ability to enact socially-appropriate behavior depends on a complex interaction between social knowledge, arousal regulation, and processing skill. Because social knowledge is formulated as a result of past experiences, which cannot be modified by environmental changes, interventions have focused on helping children manage feelings of anger and frustration, process social information accurately, and generate appropriate solutions to problems.

Successful problem solving occurs with the skillful completion of six SIP steps: a) encoding social cues; b) interpreting cues; c) formulating goals; d) generating responses; e) evaluating and selecting responses; and f) enacting a response (see *Figure 2*; Crick & Dodge, 1994; Lemerise & Arsenio, 2000). Affective processes such as emotional understanding and emotion regulation facilitate this process. Alternately, difficulty processing social information and difficulty identifying, expressing, and coping with strong emotions increases the likelihood of an aggressive response (Crick & Dodge, 1996; Crick & Werner, 1998; Eisenberg et al., 2001; Hanish & Guerra, 2002; Hubbard, 2001; Hughes, Meehan, & Cavell, 2004; Yoon, Hughes, Cavell, & Thompson, 2000).

The SIP model was originally developed in the mid-1980s (Dodge, 1986). Elaborating on D'Zurilla and Goldfried's (1971) *social problem-solving model* (SPS), it posited that children encode and interpret social cues during the problem-solving process. In the mid-1990s, the model was reformulated to better account for the role of affective processes, latent mental structures, and reciprocal effects (Crick & Dodge, 1994). Over the past twenty years, dozens of studies have focused on understanding social information processing in children (Courtney & Cohen, 1996; Milich & Dodge, 1984; Richard & Dodge, 1982; Dodge et al., 2003).

Figure 2: Crick and Dodge (1994) Reformulated SIP Model



In the reformulated model, latent mental structures and affective processes interact with a series of six SIP steps to affect behavior (Crick & Dodge, 1994). Social knowledge

and emotional arousal, for instance, affect SIP processing; and, conversely, the way in which social information is processed influences social knowledge and emotional arousal. Latent mental structures include emotional and behavioral *scripts* (i.e., automatic emotional and behavioral responses), *schemas* (i.e., mental patterns that organize and simplify complex experiences), *heuristics* (i.e., rules of thumb that guide behavior in specific social situations), as well as moral concepts such as values and beliefs. Whereas these mental structures are seen as stable, experience-based patterns of cognition that are stored in memory, social information processing is conceptualized as online (in the moment), preconscious cognitions that are more dependent upon situational and contextual contingencies (Dodge, Laird, Lochman, Zelli, & Conduct Problems Prevention Research Group, 2002). Although these knowledge structures play an important role in children's social cognitions, a discussion of latent mental structures lies beyond the scope of this paper; for a discussion, see Burks et al. (1999).

Defining Key Constructs

The crux of the SIP model lies in the articulation of six steps: (a) *encoding* (b) *interpretation*; c) *goal clarification*; d) *response search*; e) *response decision*; and f) *enactment* (see *Figure 2*). Summarized briefly, *encoding* refers to the perception of external and internal cues. External cues relate to features of the social environment, whereas internal cues relate to cognitive indicators of emotional state (e.g., negative thoughts) and physical indicators emotional arousal (e.g., heart rate, muscle tension, perspiration). *Interpretation* involves making sense of internal and external cues. Often this requires having to infer the intentions of others. Intent attributions involve making inferences about "the motives of others, based on interpretations of social cues" (Crick, Grotpeter, & Bigbee, 2002, p. 1135).

Once an attribution of intent is made, one is able to formulate a desired goal. This occurs during the *goal clarification* phase. Once a goal is formulated, one must generate an array of responses. This process of generating alternative responses is called *response search*.

In the *response decision* phase, children make means-end and outcome-based judgments – *response valuation* and *outcome expectancies* – about the solutions they generated (Fontaine, Burks, & Dodge, 2002) and make *self-efficacy evaluations* about their ability to execute various responses and assess which responses will help them to obtain desired goals. In response valuation, children evaluate the appropriateness of various responses. When children produce outcome expectancies, they anticipate the consequences of each response and decide whether they would be favorable or unfavorable. Self-efficacy evaluations involve children assessing how confident they feel about their ability to successfully execute a particular response. Consistent with a social learning model, children's confidence in enacting aggressive behavior relates to how often the child has engaged in this behavior and how frequently the behavior has been reinforced. Finally, *enactment* occurs once children act upon the response they have chosen.

Emotion regulation is another important aspect of the SIP model. Definitions for emotion regulation vary, but the term is most often used to connote an individual's ability to attenuate emotional and behavioral arousal (Eisenberg & Fabes, 1992). Walden & Smith (1997, p. 8) offer a more comprehensive definition that states: "emotion regulation processes can act to initiate, modulate, or maintain emotional arousal, experience, cognitions, and behaviors." Citing Campos, Mumme, Kermoian, & Campos (1994), they state that regulation can occur at one or more of three levels: a sensory level (physiologically), a cognitive level (information processing), and a behavioral level (coping behavior).

Behavioral regulation includes functional aspects such as behavioral self-control, the use of display rules (i.e., an age-appropriate display of affect that is consistent with social norms), and appropriate speech volume.

The Role of SIP Skills in Aggressogenic Processes

Empirical support for the *social information-processing model* (Crick and Dodge, 1994; Lemerise & Arsenio, 2000) as a framework for understanding childhood aggression has grown substantially over the past decade (Camodeca, Goossens, Schuengel, & Terwogt, 2003; Dodge, 2003; Dodge & Coie, 1987; Hughes et al., 2004; Lochman & Wells, 2002). The generalizability of the model is supported by the diversity of study samples, which represent a range of ethnic and socioeconomic backgrounds (Cortney & Cohen, 1996; Dodge et al., 2002, 2003; Hughes et al., 2004; Lochman & Dodge, 1994; Musher-Eizenman, Boxer, Danner, Dubow, Goldstein, & Heretick, 2004; Zelli, Dodge, Lochman, Laird, & CPPRG, 1999a; 1999b) and nationalities (Camodeca et al., 2003; Matthys, Cuperus, & Van Engeland, 1999; Orobio de Castro et al., 2005). Several studies have been conducted with African American samples (Hubbard, Dodge, Cillessen, & Schwartz., 2001; Schwartz et al., 1998; Shure & Spivack, 1980). Findings from these studies support similar patterns of SIP deficits in children who display above-average levels of overt aggression.

More recently, developmental psychologists have begun to investigate whether different subtypes of aggression (e.g., proactive/reactive and overt/relational) relate to particular patterns of SIP deficits (Arsenio & Lemerise, 2001; Crick & Dodge, 1996; Hubbard et al., 2001). This section will review the empirical research on SIP in relation to overt aggression and, where possible, discuss findings in relation to reactive and proactive aggression. Preliminary findings suggest that reactive aggression is more highly associated

with processing errors in the earlier phases of SIP (e.g., encoding and interpretation) and that proactive aggression is more highly associated with processing errors during the later phases – goal clarification and response decision (Arsenio & Lemerise, 2001; Smithmyer et al., 2000). During the *encoding* phase, reactively aggressive (RA) children exhibit a "social-perceptual bias," in that they tend to notice fewer cues in the environment and are more likely to encode negative cues (Courtney & Cohen, 1996; Waldman, 1996). In addition, they are more likely to lack *emotional understanding* – lacking awareness of their own feelings and misperceiving the emotions of others (Denham et al., 2002; Dodge et al., 2002). Alternately, children who are able to accurately process emotional cues in themselves and others (e.g., *emotion-processing skills*) are more likely to have higher levels of emotional understanding and a greater ability to experience empathy towards others (Schultz, Izard, & Bear, 2004).

During the *interpretation* phase, RA children tend to exhibit a "hostile attribution bias" (Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002), a tendency to assign hostile intent to ambiguous social situations (Crick et al., 2002; Matthys et al., 1999). The association between hostile intent attribution and aggression is supported by a number of studies (Orobio de Castro et al., 2002; Hudley & Graham, 1993; NICHD Early Child Care Network, 2004). In a recent longitudinal study of 1100 children from preschool-to third-grade, children belonging to the stable-high externalizing behavior trajectory were more likely than children in the other trajectories to make hostile intent attributions (NICHD Early Child Care Network, 2004). An intervention study conducted by Hudley and Graham (1993) found that aggressive children who were randomly assigned to a 12-lesson attributional intervention (n=20) had lower mean levels of hostile attribution and reactive aggression at

posttest than aggressive children in the comparison (n=22) and no-treatment (n=24) conditions.

Children generate alternative solutions during the *response search* or *response access* phase. There is some evidence to suggest that RA children generate fewer alternative solutions in response to different social situations (Richard & Dodge, 1982). A test of the mediational role of problem solving variables targeted by the *I Can Problem Solve* program supports this finding; African American, 4- to 5-year old children (N=219) who had improved ratings of behavioral adjustment (measured as a composite of impulsivity, emotionality, and overt aggression) were significantly more likely than unimproved children to exhibit improved alternative solution thinking skills (Shure & Spivack, 1980). Another study, conducted with a sample of 585 predominantly White (82% European American) elementary school children, suggests that improving response search skills may help to reduce the risk of aggression in peer-rejected children (Dodge et al., 2003). Exploring the mechanisms by which social preference in kindergarten predicted later aggression, this study found that response generation in Grade 2 explained 16% of the total effect of kindergarten social preference on Grade 3 teacher-rated aggression (Dodge et al., 2003).

Arsenio and Lemerise (2001) contend that proactively aggressive (PA) children are adept at encoding and interpreting cues, but display a bias favoring aggression during the *goal clarification* and *response decision* phases of SIP (Arsenio & Lemerise, 2001, p. 66). During the *goal clarification* phase, aggressive children tend to generate fewer goals and are less likely to conceive friendly goals (Erdley & Asher, 1996; Murphy & Eisenberg, 2002). A study conducted with 273 4^{th}-through-6^{th} grade youth found that physically aggressive youth endorsed goals of self-interest, retaliation, and social dominance (Delveaux & Daniels, 2000).

During the *response decision* phase, PA children exhibit biased processing patterns. Additional research suggests that youth who are aggressive evaluate aggressive responses in positive or normative ways (Crick & Werner, 1998; Fontaine, et al., 2002; Vernberg, Jacobs, & Hershberger, 1999; Zelli et al., 1999), whereas prosocial children do not (Nelson & Crick, 1999).

Children who are aggressive commonly hold attitudes and beliefs that favor or validate the use of aggression (Tapper & Boulton, 2004). Studies indicate that children's response valuations may be related to provocation type and gender. A study of 1,166 predominantly White third-through-sixth grade children found that girls had more favorable evaluations of relational aggression, whereas boys had more favorable evaluations of overt aggression (Crick & Werner, 1998). In contrast, in a study examining children's responses to hypothetical scenarios, conducted with an ethnically-diverse sample of 387 third- through fifth- graders (approximately 50% White and 50% Black), response valuations were not moderated by gender or ethnicity (Zelli et al., 1999).

Research on children's self-efficacy evaluations suggests that aggressive children are more confident about their ability to enact aggressive responses than they are about their ability to enact prosocial responses (Erdley & Asher, 1996; Matthys et al., 1999). Finally, aggressive youth, especially boys, are more likely to believe that aggressive behavior will lead to favorable outcomes (Cuddy & Frame, 1991; Schwartz et al., 1998). Indeed, in a recent study conducted with preadolescent youth, outcome expectations were found to partially mediate the intervention effects of the *Coping Power Program (*Lochman & Wells, 2002) on later delinquency, measured one year after intervention completion, when boys had just completed the sixth or seventh grade).

Indeed, a variety of potential aggressogenic risk mechanisms may be derived the SIP model. As this review has suggested, different SIP skills may be implicated in the development of different types of aggression (proactive/reactive; overt/relational). For instance, faulty encoding and interpretation and accompanying feelings of anger appear to trigger the use of reactive aggression. Alternately, setting a social dominance goal and having positive outcome expectations for aggression appear to be associated with proactive aggression.

Studies that attempt to differentiate the relationship between different SIP skills and problem behavior provide more useful knowledge for intervention. For example, if the above examples were true, then one would expect (a) interventions altering children's aggression-related attitudes and beliefs to reduce proactive aggression, and (b) interventions seeking to reduce emotional reactivity and hostile intent attributions to reduce reactive aggression. However, even this more careful approach to intervention design does not involve enough complexity. The ways in which multiple factors, such as race/ethnicity, SES, gender, culture, and family context, moderate risk mechanisms must be tested, in order to ensure that this model can be translated properly in the design and development of interventions for aggressive youth.

The Role of Emotion Regulation in Aggressogenic Processes

Emotion regulation difficulties have been linked to physically aggressive behavior (see Eisenberg & Fabes, 1999, for a review). In response to emotional arousal, children with low emotion regulation are more likely to rely on automatic schema and scripts, perceive fewer social cues, generate fewer solutions, and select aggressive responses (Bierman et al., 1993; Eisenberg et al., 2001; Izard, 2002). In addition, children with emotion regulation

difficulties are more likely to display strong negative (or positive) affect, which makes them more likely to induce negative peer responses (Eisenberg & Fabes, 1992; Hubbard, 2001), more prone to peer rejection and victimization (Schwartz et al., 1999), and more likely to experience negative adjustment outcomes (Lengua, 2003). Conversely, children who are able to understand and manage feelings are more likely to demonstrate skill in solving social problems, display appropriate levels of affect, enact prosocial behavior, and induce positive peer responses (Lemerise & Arsenio, 2000; Rydell, Berlin, & Bohlin, 2003).

SIP Skills and the Role of Emotion

Although depicted in only two steps of the SIP model (i.e., *encoding* and *goal clarification*), affective processes are theorized to impact all phases of social information processing. In their 1994 review (pp. 81-82), Crick and Dodge (1994) assert that affective- and social information-processing components of the model exert reciprocal influences on each other. For example, one's style of emotional responding (i.e., *emotionality*) is posited to influence intent attribution and goal selection. Unregulated emotional arousal is posited to influence the encoding and interpretation of external cues, in some cases causing a pattern of *preemptive processing* – where the brain relies on internal mental scripts and schemas to process information rather than conscious, rational decision-making skills (Burks et al., 1999). Negative intent attribution and goal selection are thought also to affect emotional responses (e.g., feelings of anger, distress, or sadness). In the response search phase, a child may experience an emotional response while generating potential responses to a social situation (the example offered by Crick and Dodge is that a victimized child who considers the response 'kick my bully in the gut' could experience some sense of emotional relief). Finally, evidence suggests that children often consider the opinions and emotions of others

when evaluating potential ways to respond to a social situation and weighing the consequences of their behavior.

Psychologists are beginning to investigate the interaction between SIP skills and emotional processes (Lochman & Wells, 2002; Orobio de Castro, Bosch, Veerman, & Koops, 2003; Orobio de Castro, et al., 2005). For example, findings from a study utilizing delay prompts to assess the effects of emotion regulation on anger and intent attribution suggest that children's intent attributions may be influenced by emotion (Orobio de Castro et al., 2003). Aggressive boys who were asked to monitor their feelings and generate emotion regulation strategies prior to responding to a hypothetical vignette involving a peer provocation had lower rates of biased intent attribution than aggressive boys in the comparison conditions. Murphy and Eisenberg (2002) conducted a study of social goals and emotionality, with a predominantly White sample of 118, 7- to 11-year-old children. During individual interviews, children were asked to think about three actual conflicts with peers they had recently had and then were asked a series of questions related to each conflict. Aggressive children were more likely than non-aggressive children to experience anger and set hostile social goals (e.g., 'I wanted to annoy him'), controlling for child's age, the gender of the peer involved in the conflict, friendship with peer involved in the conflict, and the nature of the provoking event. Nonaggressive children reported less anger and were more likely to set friendly social goals (e.g., 'I wanted her to be happy').

Only a handful of studies have attempted to analyze the interplay between these variables in relation to aggression. Three studies examining the relationships between SIP skills, emotion, and aggression were identified in this review. First, an evaluation of the *Attributional Program* (Hudley & Graham, 1993) offered support for an *Attribution-*

Emotion-Action model of aggression. According to this model, biased intent attributions produce feelings of anger which, in turn, trigger an aggressive response (Graham et al., 1992). At posttest, students who participated in an intervention designed to decrease hostile attribution were rated by teachers as displaying less anger than non-program participants [$t(19)=5.75$, $p<.001$]. In addition, participants with lower levels of hostile attribution also had lower ratings of anger and aggression. Second, Schultz et al. (2004) sought to assess whether encoding skills related to emotion processing (the encoding of emotions in others) explained the relationship between anger and aggression. The study was conducted with a predominantly White sample of 182 first- and second-grade children. Findings indicated direct effects between anger and aggression and direct effects between emotion processing skill and aggression, but did not support a significant indirect effect between anger and aggression via emotion processing skill. Third, Musher-Eizenman and colleagues (2004) tested the relationships between response decision, anger control, and aggression, with 778 fourth-through-sixth grade children, from nine schools – four urban and five suburban. Principals from each school randomly selected six classrooms (two classrooms per grade) to participate in the study. This study found that retaliation approval beliefs mediated the relations between anger control and aggressive behavior but not between impulsivity and aggression. One possible implication of this finding is that altering attitudes and beliefs may not an effective tactic for reducing reactive aggression.

To date, the relationship between emotion regulation and SIP skill remains unclear. Further research examining the way in which SIP-skill deficits and emotion regulation difficulties interact to encourage the enactment of aggressive behavior is needed.

SIP Skills, Emotion Regulation, and Gender

The literature suggests that, on average, girls and boys process and respond to social information in different ways (Crick, Grotpeter, & Bigbee, 2002; Murphy & Eisenberg, 2002). For example, girls and boys' social cognitions have been found to differ by provocation type. Girls tend to rate direct aggression less favorably than boys and rate indirect aggression more favorably (Crick & Werner, 1998). At the same time, girls rate relational victimization as being more hurtful and hostile than boys (Crick, 1995). Girls also tend to report friendlier goals and response selection than boys (Murphy & Eisenberg, 2002). In addition, studies also suggest that girls have higher levels of emotion regulation than boys (Gross & John, 2003). Eisenberg et al. (2001) stated that improving emotion regulatory processes may do more to promote behavioral improvement for boys than for girls. In light of gender differences in SIP skill and emotion regulation, and preliminary evidence to suggest that relationships between emotion and aggression to vary by gender, greater exploration of the SIP model as a framework for aggression in females is needed.

Evaluating the SIP Model

Utilizing the SIP model as a conceptual framework to guide intervention has both strengths and drawbacks, but two major strengths are clear. First, the association between SIP-skill deficits and overt aggression has strong empirical support with diverse samples. Physically and verbally aggressive children consistently demonstrate deficits (and/or biases) in SIP skill in studies conducted with normative and clinical samples of different racial/ethnic backgrounds, socioeconomic classes, and nationalities. Second, measures assessing social-information processing skill, such as the Home Interview with Child (Dodge, Bates, & Pettit, 1990), the Social Problem Solving Inventory (SPSI; D'Zurilla & Nezu, 1990), and the Social

Problem Solving Assessment Measure – Revised (SPSAM-R; Butler & Meichenbaum, 1981), appear to be valid and reliable across diverse populations.

Several limitations of this framework must also be acknowledged. One problem concerns the measurement of SIP skill. Unlike measures of social competence – which can be rated by parents, teachers, and peers – measures of social cognition primarily rely on child report. Therefore, a multi-informant approach to assessing SIP skill is often not possible. This may result in issues relating to the validity of the measurement. In addition, SIP measures are often administered by trained staff members. As a result, variables associated with the administration of the measure (e.g., videotaped versus audiotaped vignettes; individual versus group interviews) could introduce some degree of measurement error. Moreover, SIP measures have not been validated with English-as-a-Second Language (ESL) students. As a result, language comprehension difficulties could affect the interpretation of situations presented in the vignettes and the meaning of response items.

A second problem relates to the breadth and complexity of the SIP model. Designing interventions requires a clear understanding of causal relationships between targeted skills and desired outcomes and of the mediators and moderators that explain or influence these relationships. Interventions with conceptual roots in the SIP model have operated under the empirically-valid assumption that each SIP skill has direct associations with aggression. . Because little is known about the relative importance of SIP skills, the current working assumption is that interventions targeting a greater number of SIP skills would be more effective at reducing aggression than those targeting a fewer number of skills. More research is needed to determine whether certain skills uniquely predict aggression (above and beyond the contribution of other skills) and examine whether SIP skills relate to aggression indirectly

via critical mediators. Also, further examination of the interrelationships between SIP skills is needed. For example, few studies have tested whether there are reciprocal relationships between SIP skills (e.g., between encoding and interpretation), as posited by the SIP model.

Moreover, few studies have examined the relationship between SIP skills and other components of the SIP model, such as arousal regulation and latent mental structures. Those that have tested these relationships often test bivariate relationships, rather than conducting multiple regression analyses to estimate the relationship between two variables, while controlling for a third variable (Camodeca & Goosens, 2004; Crick et al., 2002; Dearing et al., 2002; Denham et al., 2002; Eisenberg et al., 2001). Lacking multivariate analyses that include both SIP and emotion-related variables as predictors, it is difficult to discern whether SIP and emotion have unique effects on overt aggression. Examining the systems of relationships between SIP, emotion, and social knowledge and between different SIP skills (and how they relate to aggression) would facilitate the identification of SIP-related risk mechanisms specific enough to inform the design of preventive interventions.

Third, few studies have attempted to examine the mediating effect of SIP skills on developmental trajectories or antisocial behavior. For instance, a study by Dodge and his colleagues (2003) suggests that response generation (i.e., the number of responses children can generate to solve a social problem) may mediate the negative effects of childhood peer rejection on later externalizing problems (Dodge, Lansford, Burks, Bates, Pettit, Fontaine, & Price, 2003). Until more analyses like these are implemented, the specific mechanisms by which SIP skills enable or deter the development of aggressive behavior will remain unclear.

A fourth concern relates to the under-representation of girls in studies examining the association between SIP and aggression (Hughes et al., 2004). Indeed, much of the empirical

support for the SIP model derives from studies conducted with all-male samples of elementary school children (Bierman, Smoot, & Aumiller, 1993; Courtney & Cohen, 1996; Cuddy & Frame, 1991; Hubbard et al., 2001; Lochman & Dodge, 1994, 1998; Matthys et al., 1999; Orobio de Castro et al., 2005; Schwartz et al., 1998; Trachtenberg & Viken, 1994; Waldman, 1996). Although several studies on mixed-sex samples have supported the role of SIP in females (Crick & Dodge, 1996; Dodge et al., 2002, 2003; Zelli et al., 1999), a greater number of exploratory studies examining the etiology of overt aggression (and the causal role of SIP deficits) in all-female samples are needed.

Finally, given preliminary evidence that certain SIP skills are more predictive of aggression in boys than in girls (Musher-Eizenman et al., 2004; Schultz et al., 2004), more research on the moderating role of gender is needed. Moderating effects for gender have been reported in several studies. A study by Schultz et al. (2004) found that emotion-processing variables (i.e., emotion attribution accuracy, anger attribution bias, and empathy) were more strongly associated with teacher-rated overt aggression for girls than for boys. Musher-Eizenman et al. (2004), in a study of social-cognitive mediators of aggression (with 778 fourth-through-sixth grade children), found that retaliation approval beliefs were stronger mediators for girls and self-efficacy evaluations were stronger mediators for boys. In contrast, a study conducted with an adolescent sample (N=124) found that favorable response valuations for aggression were more predictive of concurrent and subsequent externalizing problems for males than for females (Fontaine et al., 2002); these contradictory findings could relate to age differences in the study samples. In light of this research, it is critical that future studies test whether aggressogenic SIP processes vary by gender, age, and risk status. To be effective at promoting positive adjustment, interventions must address risk

mechanisms that are not only specific to the problem but also relevant to the target population (Coie et al., 1993).

Conclusion

Despite the wealth of support for the SIP model, the majority of social-cognitive interventions for aggressive behavior are based on a conventional social problem solving (SPS) model that involves: a) problem definition and formulation, b) generating alternative solutions, c) decision-making based on means-end thinking, and d) solution implementation and verification (D'Zurilla & Goldfried, 1971; Shure & Spivack, 1980). The SIP model includes these SPS skills and adds at least six additional components (i.e., encoding, interpretation, goal clarification, and response valuation, self-efficacy evaluation, and arousal regulation) – each of which appear to have clear links to aggression (Murphy & Eisenberg, 2002; Crick & Werner, 1998; Eisenberg et al., 2005; Shultz et al., 2004). Interventions with a stronger base in the SIP model may have greater potential to decrease aggression and modify trajectories of antisocial behavior. The following chapter reviews such interventions, in order to evaluate their overall effectiveness and identify strengths and limitations.

Chapter III

School-based Violence Prevention with Social-Cognitive Foundations

Considering the evidence relating SIP to aggression in childhood, one would expect elementary school-based violence-prevention programs with social-cognitive foundations to have positive effects on aggressive and disruptive classroom behavior. This chapter reviews universal and indicated interventions informed by emotion-integrated, SIP-related models and evaluates effects on social-cognitive, emotional, and behavioral outcomes. Mediating and moderating effects are described. Three key questions guide this review: a) do programs affect theoretical mediators (social-cognitive and emotional skills); b) do programs reduce aggression; c) are effects moderated by gender; and d) do social-cognitive and emotional processes explain effects on aggression? Following this review of study findings, study limitations and implications for future research will be discussed

According to the Institute of Medicine (IOM) classifications, universal programs are activities targeted to the general population. Alternately, indicated programs are activities "targeted to individuals in high-risk environments, identified as having minimal but detectable signs or symptoms foreshadowing disorder" (Mrazek & Haggerty, 1994). In indicated violence prevention research, children are typically selected on the basis of having above-average levels of teacher-rated aggressive or disruptive behavior.

Two universal (*Promoting Alternative Thinking Strategies* and *Second Step*) and two indicated (*Anger Coping Program* and *Attributional Program)* interventions are reviewed (see *Table 1*). These programs were selected because they (a) utilize an SIP-related, social-

Table 1: Social-cognitive Interventions

Program	Sample	Program Content	Duration/ Intensity
PATHS (Promoting Alternate Thinking Strategies) *Universal	Ages 5-12; regular ed., special ed., & hearing-impaired; boys and girls	Targeted skills: a) self-control; b) emotional understanding; c) verbal and nonverbal communication skills; and d) problem-solving skills. Children learn to identify and label feelings, engage in perspective-taking, monitor and manage emotions, and express feelings appropriately to others. Children receive instruction on how to cope with stress, control impulses, and manage emotions. SIP-related: Encoding of internal and external cues. Emotion-related: Emotional understanding, emotional expressiveness, self-regulation.	Duration: 5 years Service Intensity: 60 lessons; 20-30 min/lesson; 3x/week
Second Step Program *Universal	Ages 4-14; boys and girls	Targeted skills: a) empathy; b) impulse control; c) problem solving; and d) anger management. Children learn how to identify and understand their own and others' emotions. They also learn how to choose positive goals, control impulsivity, manage emotional arousal, and evaluate consequences of behavior. SIP-related: Encoding of internal cues; goal clarification; and response decision. Emotion-related: Impulse and anger management.	Duration: 16-20 weeks Service Intensity: 30-32 lessons; 35 min/lesson; 1-2x/week
Attribution-al Program *Indicated	Ages 8-12; boys only	Targeted skills: a) search for clues; b) interpret clues; c) generate possible attributions; and d) generate decision rules for behavior in ambiguous situations. Children identify intentions, generate possible causes in situations where intent is ambiguous, and generate decision rules about how to respond in ambiguous situations. SIP-related: Encoding external emotional cues; interpretation of cues; response search. Emotion-related: None.	Duration: 1 year Service Intensity: 12 lessons
Anger Coping Program *Indicated	Ages 8-12; boys only	Targeted skills: a) emotion regulation; b) emotional understanding; c) perspective-taking; d) alternative solution thinking; and e) means-end thinking. SIP-related: Encoding internal cues; response search; and response decision - outcome expectancy component. Emotion-related: Emotional awareness and understanding; anger management.	Duration: 4-5 months Service Intensity: 12-18 sessions; 45-60 min/session; 1x/week

problem solving framework; b) are school-based; and c) are classroom-focused. Programs were not considered if they: a) included training unrelated to social problem solving and emotion – such as friendship skills, tutoring, or mentoring; b) targeted multiple social domains; c) had never been tested with elementary-school children; and d) had never been evaluated with at least one randomized, controlled trial. Curriculum content, sample characteristics, study design, and service duration and intensity for these programs are noted in Appendix A.

Sixteen studies measuring the impact of these programs were identified and evaluated. Seven out of sixteen studies utilized randomized controlled designs and the majority of the remaining studies utilized quasi-experimental designs. Follow-up data collection ranged from three months to two years after intervention completion. Main effects on posttest and follow-up outcomes are summarized for each program, and, where possible, mediating effects of targeted skills and moderating effects of gender and risk status are noted (see Appendix A). Reported findings are limited to social-cognitive, emotional, and behavioral outcomes, and effect sizes are noted wherever possible. Appendix B contains a summary of findings for randomized, controlled studies.

Do School-based Programs Modify Social-Emotional Skills and Reduce Aggression?

By and large, the programs reviewed were effective at promoting social-emotional competence and decreasing aggression in both indicated and universal samples (see Appendix A). The following section will review social-cognitive, emotional, and behavioral effects for each program.

Social-Cognitive Outcomes

All four programs measured social-cognitive outcomes. Measured outcomes included SIP-related outcomes such as response search, outcome expectancies, self-efficacy evaluation, response selection, and attitudes toward aggression, as well as non-SIP outcomes such self-esteem and knowledge and skill acquisition.

SIP-related skills were measured primarily by presenting children with hypothetical vignettes and then asking them to respond to a series of questions. Vignettes were most often read aloud to students and accompanied by illustrations. For example, Greenberg and Kusché (1998) used the *Social Problem Solving Assessment Measure – Revised (SPSAM-R;* Elias, Larcen, Zlotlow, & Chinsky, 1978) to assess social and emotional understanding in 32 hearing-impaired children. This instrument consists of six written stories which describe three different types social situations that elementary school children commonly encounter (i.e., wanting something another peer has; being unjustly blamed for misbehavior; and being excluded by a group of peers). The stories were presented on story cards and accompanied by illustrations; research staff signed the stories to the children and asked a series of questions related to perspective-taking (e.g., "How do you think X is feeling?"), response generation (e.g., "What could X do?), outcome expectancies (e.g., "What do you think will happen next?"), and means-end thinking (e.g., "What happened here?").

Hudley and Graham (1993) examined participants' responses to a staff-administered questionnaire, involving five hypothetical vignettes that described peer provocations of varying intents (e.g.. prosocial, accidental, ambiguous, or hostile). After listening to each vignette, participants were asked to answer six multiple choice questions relating to intent attribution (e.g., "Do you think he did this on purpose") and emotional response (e.g., How

angry would you feel if this happened to you?). To measure response selection, they were then asked 'What would you do if this happened to you?' and were told to select one out of six possible responses.

Effects on social cognitive outcomes are discussed in this section. Study findings indicated that the interventions had varied effects on SIP-related skills (see Appendix A for a brief overview of these effects.

Universal programs. The effects of the *PATHS* program on social problem-solving skills were reported in two studies (Greenberg & Kusché, 1998; Kam, Greenberg, & Kusché, 2004). Kam et al. 2004, in a study of 133 first- to third-grade students with disabilities, studied program effects on response search and response decision skills (i.e., outcome expectancy, self-efficacy evaluation, and response selection). Marginal effects were found for response search skills (i.e., the treatment group had a 22.5% reduction in aggressive solutions and control group had a 15.5% reduction in aggressive solutions); and no significant changes in outcome expectancies and self-efficacy evaluations were found.

Greenberg and Kusché (1998), in their study of 57 elementary school children with profound hearing loss (83% European American; 17% other; ages 5 to 12), obtained significant treatment effects for SIP outcomes at posttest. Findings suggested increased perspective-taking ($F(1,52)= 5.7$, $p<.05$), outcome expectancies ($F(1,52)= 9.7$, $p<.001$), means-end thinking ($F(1,52)= 20.1$, $p<.001$), response generation ($F(1,52)= 8.1$, $p<.05$), and generation of prosocial solutions ($F(1,52)= 34.0$, $p<.001$). Decreases in the generation of neutral ($F(1,52)= 4.4$, $p<.05$) and negative solutions ($F(1,52)= 4.9$, $p<.05$) were also found.

A quasi-experimental evaluation of the *Second Step* program (Orpinas et al., 1995) found that intervention students demonstrated a significant improvement ($F(2,205)=4.22$,

$p<.05$) in response valuation (measured as "attitudes toward provoked violence") and a marginally significant increase ($F(2,221)=2.71$, $p<.10$) in violence-related knowledge and skills. Another study of this program conducted with 714 preadolescents (grades six through eight) found significant reductions in aggression, with an omnibus multivariate test indicating a significant time by group interactions for physical aggression ($F(1,252)=37.03$, $p<.001$, ES=77), verbal derogation ($F(1,252)=26.42$, $p<.001$, ES=65, and social exclusion ($F(1,252)=23.36$, $p<.001$, ES=61). The group variable was a dichotomous variable indicating whether a child was in his or her first or second year of school. Effect sizes for participants in the Year 2 group were stronger than those for participants in the Year 1 group. Compared to Year 2 controls, Year 2 program participants had a significantly lower endorsement of physical aggression ($t(291)=-4.29$, $p<.001$; ES=-.50), verbal derogation ($t(289)=-5.07$, $p<.001$; ES=-.60), and social exclusion ($t(292)=-6.29$, $p<.001$; ES=-.73). (Van Schoiack-Edstrom, Frey, & Beland, 2002).

Indicated programs. The two indicated programs also yielded mixed results. The *Anger Coping Program (ACP)* had effects on boys' self-esteem and perceived social competence (Lochman & Curry, 1986; Lochman et al., 1984; Lochman et al., 1989; Lochman, 1992) but did not yield consistent effects on problem solving skill; that is, Lochman et al. (1984) detected no effects on generating alternative solutions, whereas Lochman (1992) did find significant effects. Boys in the *Attributional Program* had fewer hostile perceptions of intentionality – $t(19)=8.08$, $p<.001$ – and less endorsement of hostile behavior – $t(19)=3.01$, $p<.05$ – than boys in the two other conditions (Hudley & Graham, 1993). This finding indicates that intent attributions and response valuations of aggression are malleable targets of intervention.

Emotional Outcomes

Three out of four programs examined emotional outcomes. Significant effects of outcomes such as emotional expressiveness (e.g., feelings vocabulary), emotional understanding (e.g., empathy), mood (e.g., self-reported depression or anger), and emotion regulation were reported (see Appendix A). Emotional outcomes were primarily measured through child self-report questionnaires, such as the Kusché Emotional Inventory (KEI; Kusché, 1984), and child interviews (Greenberg et al., 1995); in one study, a teacher-rated instrument was used (Greenberg & Kusché, 1998).

Universal Programs. Most findings relating to emotional arousal and emotion processing skill were reported by studies conducted on universal programs. Evaluations of the *PATHS* program found the program led to increased feelings vocabulary and emotional understanding (Greenberg et al., 1995), improved emotional recognition skills (Greenberg & Kusché, 1998), decreased self-reported depression, and increased negative feelings vocabulary (Kam et al., 2004).

In a randomized, controlled evaluation of the *PATHS* program, Greenberg et al. (1995) reported the emotional outcomes of second- and third-grade students (N=286; 53% European American, 32% African American, and 10% other). This study found students in the *PATHS* program had an increased vocabulary for positive emotion words ($F(1,282)= 21.5$, $p<.001$) and negative emotion words ($F(1,282)= 49.9$, $p<.001$). Although there were no increases in emotional awareness, *PATHS* participants had greater emotional understanding (i.e., knowing how others feel) ($F(1,268)= 7.5$, $p<.01$) and greater emotional control ($F(1,280)= 33.1$, $p<.001$) than students in the comparison group at posttest. Another evaluation of the *PATHS* program (Kam et al., 2004) found a medium intervention effect for

negative feelings vocabulary two years after program completion, in a predominantly African American, low-income sample (N=350) of six- to eight-year-old students with disabilities (t= 2.83, p<.05; Cohen's d = .54). This study also found small to moderate effects on teacher- and child-rated internalizing behavior (t=2.48, p<.05; Cohen's d=.22) and depression (t = 3.13, p<.05; Cohen's d = .49).

The *Second Step* program also promoted emotional understanding. A study of this program was conducted with a predominantly African American sample (N=149; 64% female; ages 11 to 14), recruited from two schools (McMahon & Washburn, 2003). No comparison group was used. A repeated-measures ANOVA found significant main effects for time (from pretest to posttest) on self-reported empathy (F(1,90)= 4.13, p<.05) , controlling for gender, grade level, and school. A significant time by school interaction was also found (F(1,90)= 6.69, p<.05), with students in School B having larger average effects than students in School A.

Indicated Programs. Out of the two indicated programs, only one program, the *Attributional Program*, studied emotion-related outcomes. This program was evaluated in one study, conducted with 101 9- to 11-year-old African American boys. The study found that aggressive intervention participants had significantly lower self-reported anger at posttest compared to pretest (t(19)= 5.75, p<.001), and aggressive boys in the other two conditions did not (Hudley & Graham, 1993).

Behavioral Outcomes

All programs appeared to reduce overt aggression, although less reliable effects were obtained for universal programs. Study findings related to the moderating effect of risk status at pretest are reviewed in the subsequent section. Measured outcomes included items related

to physical aggression, disruptive behavior, externalizing behavior, prosocial behavior, and social competence (see Appendix B). Study outcomes were assessed by a variety of informants and methods – i.e., teacher- and parent-rated surveys, peer nominations, and behavioral observations.

Universal programs. A study of the *PATHS* program with 57 children with profound hearing loss from 11 classrooms (Greenberg & Kusché, 1998) found intervention students to have significant improvements on teacher-rated emotional adjustment ($F(1,53)= 5.1$, $p < .05$; Cohen's $d = .25$) and on parent-rated social competence ($F(1,41)=4.5$, $p<.05$; Cohen's $d = .71$), but not for externalizing or internalizing symptoms. A more recent study of the *PATHS* program, delivered to 133 children from 7 schools in the Northwest region of the U.S., examined posttest effects utilizing individual growth curve analysis (Raudenbush & Bryk, 2002). Level 1 (time) included four observations per child; level 2 (child) contained 133 subjects. A small program effect was found on the rate of teacher-rated externalizing behavior on the *Teacher Report Form* (*TRF*; $t= 2.03$, $p<.05$; Cohen's $d=.18$); intervention participants had lower rates of externalizing behavior than non-participants (Kam et al., 2004).

In the only randomized, controlled study of the *Second Step* program (Grossman et al., 1997), intervention students were observed to have greater decreases in physical aggression and greater increases in neutral/prosocial behavior at posttest in classroom, playground, and cafeteria settings than controls. At posttest, overall scores for physical aggression (i.e., reflecting a combined score for classroom, playground, and cafeteria settings) decreased at a rate of .46 events per hour ($p<.05$) and neutral/prosocial behavior

increased at a rate of 3.96 events per hour ($p<.05$). Effects for observer-rated, classroom physical aggression were maintained at the six-month follow-up evaluation.

Effects were only obtained for observer-rated (and not parent- or teacher-rated) aggression. Observers were trained research staff and were not blind to treatment assignment. Behavioral observations were conducted on a random selection of 12 children from each participating classroom. A total of 60 minutes of behavioral observation was conducted on each student with 10% of the observations conducted by two raters. The interrater reliability was strong for prosocial/neutral behavior ($\kappa = .92$) and fair for physical behavior ($\kappa = .50$).

Three additional evaluations of the *Second Step* program reported effects on behavior – one study with 109 preschool children (McMahon, Washburn, Felix, Yaking, & Childrey, 2000), another study with 70 elementary-age children (Taub, 2001), one study with 223 pre-adolescent children (Orpinas, Parcel, McAlister, & Frankowski, 1995). McMahon et al. (2000) found no effect on teacher-rated aggression, but did find a significant effect for time (time 2, and time 3) on three types of observer-rated aggression (i.e., verbal aggression, physical aggression, and disruptive behavior). However, the findings of this study are weakened by the use of a time-series sample design with no comparison group. Taub (2001) found the reverse to be true, finding significant effects at posttest for teacher-rated antisocial behavior and social competence and failing to find observer-rated effects for five different types of problem behavior in the classroom. Finally, Orpinas et al. (1995) found significantly lower levels of physical aggression for boys in two out of six intervention classrooms, but not for the intervention group overall and not for intervention group girls.

Indicated programs. Lochman et al. (1984) randomly assigned 76, 9- to 12-year-old teacher-referred boys (53% African American and 47% Caucasian) to one of four conditions:

Anger Coping Program (*ACP),* Goal Setting *(GS),* Anger Coping Plus Goal Setting *(ACGS),* or a no-treatment control group. Boys randomly assigned to the Anger Coping conditions had less observer-rated, disruptive off-task behavior – $F(1,72) = 5.37$, $p<.05$ – and less parent-rated aggression – $F(1,68) = 6.42$, $p<.05$ – than boys in the non-Anger Coping conditions (e.g., goal setting and control group) over time. Lochman et al. (1989) replicated this finding with another high-risk sample (n=32) of boys. Boys in the *ACP* displayed lower levels of disruptive off-task behavior – $U(11,8)=22.0$, $p<.05$ – and marginally lower levels of parent-rated aggression – $U(11,8)=25.0$, $p<.10$ – than untreated children (Lochman et al.,1989). These findings are mitigated, however, by the small sample sizes in each study and the lack of random assignment in the second study. In addition, because the principal investigator (and program developer) was involved in all evaluations of this program, the influence of investigator bias must be considered.

An evaluation of the *Attributional Program (AP)* found that, compared to aggressive boys in the two other conditions (i.e., attribution training: n=22; control: n=24), aggressive boys in the *AP* condition (n=20) had: (a) less observed negative verbal behavior in a peer provocation task – $F(2,64)=5.01$, $p<.01$; and (b) less teacher-rated reactive aggression – $F(2, 126) = 3.76$, $p<.05$ (Hudley & Graham, 1993). The authors hypothesized that the positive findings of the peer provocation task suggest generalization to actual situations involving ambiguous peer provocations. Because standard deviations were not provided, effect sizes for these outcomes could not be estimated.

Mediating and Moderating Effects

Six out of 16 studies tested mediating and moderating effects. Four studies examined moderating effects of gender and factors relating to risk status, such as pretest scores on

aggression and depression, and two studies tested the mediating effects of social-cognitive and emotional skills such as anger, hostile attribution, and empathy.

Program-by-gender moderating effects were tested in evaluations of the universal programs, *PATHS* and *Second Step*. In an evaluation of the *PATHS* program, Greenberg et al. (1995) conducted a four-way, repeated-measures ANOVA, with three between-subjects factors (Intervention Status, Educational Placement, and Gender) and one within-subject factor (Time). No significant interactions were found. Gender moderated the effect of the *Second Step* program in two studies conducted with preadolescent youth (Orpinas, Parcel, McAlister, & Frankowski, 1995; Van Schoiack-Edstrom et al., 2002). This program emphasized the development of emotional understanding and impulse control. In the first study (n=223), the program reduced boys' self-reported aggression, but had no effect on girls' aggression (Orpinas et al., 1995); in the second study (n=714), the program improved girls' attitudes towards aggression, but did not improve boys' attitudes (Van Schoiack-Edstrom et al., 2002).

Finally, risk status moderated the effects of the *PATHS* and the *Attributional Program*. The *PATHS* program, implemented with 286 7- to 9-year-old youth, was found to have a protective effect for two subgroups of children – those with special needs and those rated by teachers as having high levels of internalizing problems at pretest. Special education students participating in *PATHS* demonstrated better understanding of the strategies others use to mask feelings than those who did not receive the intervention. Children with high levels of internalizing problems made the greatest relative improvements from pretest to posttest on emotional awareness (Greenberg et al., 1995). Pretest levels of aggression also moderated the impact of the *Attributional Program* (Hudley & Graham, 1993); the program resulted in a

smaller incidence of negative judgments in response to ambiguous scenarios in aggressive participants, compared to their non-aggressive counterparts. In addition, teacher-ratings of aggressive behavior improved significantly for aggressive participants but not for non-aggressive participants. Moderating effects for race/ethnicity were not reported in the studies reviewed.

Only 2 out of 16 studies reported mediating effects (Hudley & Graham, 1993; McMahon & Washburn, 2003). Both evaluations relied on multiple regression techniques to test mediation. McMahon and Washburn (2003) found that pretest-to-posttest increases in empathy were associated with lower self-reported aggression at posttest (R^2=.22, F(2,91)= 13.07, p<.05). Hudley and Graham (1993), using univariate and stepdown analyses, found that the program effects on anger were partially mediated by hostile intent attribution. Anger and intent attribution mediated program effects on aggression.

Discussion and Limitations of Intervention Research Findings

Overt aggression in childhood is a major precursor for later maladjustment (Patterson, Forgatch, Yoerger, & Stoolmiller, 1998; Shaeffer et al., 2003; Stormshak, Bierman, & The CPPRG, 1998). Social-emotional deficits (in SIP skill, emotional regulation, and emotion processing skill) relate strongly to aggressive behavior in children. Universal and indicated school-based programs targeting social-emotional skills appear to promote social-cognitive, emotional, and behavioral improvements in elementary school youth. The current chapter reported results from four elementary-school based programs informed by emotion- and SIP-related conceptual frameworks. The majority of studies utilized multiple measures and informants to assess program outcomes. The following discussion summarizes treatment effects and identifies some drawbacks of the intervention studies reviewed.

Summary of Program Effects

Social-cognitive-emotional outcomes. Understandably, given the variability of SIP-related content in program curricula (see *Table 1*), interventions had varied effects on social-cognitive outcomes. Marginal and significant effects were found for encoding external clues (Greenberg & Kusché, 1998), hostile attribution (Hudley & Graham, 1993), response valuation (Hudley & Graham, 1993; Orpinas et al., 1995; Van Schoiack-Edstrom et al., 2002), outcome expectancies (Kam et al., 2004), social self-efficacy (Lochman et al., 1989), self-esteem (Lochman et al., 1984) and response selection (Greenberg & Kusché, 1998; Kam et al., 2004). They also promoted emotional competence, by improving emotional understanding – e.g., emotional labeling, emotional recognition, and empathy – (Greenberg et al., 1995; McMahon & Washburn, 2003), facilitating emotional expression (Kam et al., 2004), and decreasing anger (Hudley & Graham, 1993) and sadness (Kam et al., 2004).

Behavioral outcomes. Finally, all interventions promoted behavioral outcomes. Intervention groups compared favorably to comparison groups in the following areas: social competence (Greenberg & Kusché, 1998), physical aggression (Grossman et al., 1997), classroom disruptive behavior (Lochman et al., 1984) and negative verbal behavior (Hudley & Graham, 1993). Universal programs appeared to have less consistent effects on aggressive behavior than indicated programs. For instance, although several evaluations of the *Second Step* program resulted in decreased aggression, these decreases were often not validated by other measures (Grossman et al., 1997; McMahon et al., 2000; Taub, 2001) or not applicable to a subgroup of children (Orpinas et al., 2000). Three out of four evaluations of the *PATHS* program examined effects on externalizing behavior. Out of these, one study found no effects (Greenberg & Kusché, 1998) and two detected small effects (Kam et al., 2003, 2004).

Conversely, nearly all evaluations of the indicated programs (i.e., *Attributional Program* and *Anger Coping Program*) resulted in decreased aggression, and findings were cross-validated by multiple informants (i.e., observers, parents, and teachers).

Limitations of Reviewed Studies

The sixteen intervention studies reviewed have important methodological and substantive limitations. First, several of the studies reviewed were affected by selection bias. Over half of the studies reviewed (10 out of 16) employed non-randomized study designs (Greenberg & Kusché, 1998; Grossman et al., 1997; Lochman & Lampron, 1988). When randomization does not occur, selection bias is likely to affect study findings (Heckman, Ichimura, Smith, & Todd, 1996). Selection bias limits the internal validity of the study by producing differences between groups that are not due to chance (Morgan, Glimer, & Harmon, 2000). Even when random assignment does occur, selection bias may affect study findings. For example, selection effects may have affected the findings of a randomized controlled study of the *Second Step* program. In this study, only 790 (66%) of the 1100 students who had been randomized to receive the program had the necessary consents and parent data at pretest (Grossman et al., 1997). Thus, study attrition may have affected equivalence between study conditions.

Second, most studies failed to employ analytic methods that account for shared variation resulting from the clustering at the classroom level. Ignoring nested effects leads to decreased standard errors for parameter estimates and an increased likelihood of obtaining significant effects (for a discussion, see Shrout & Bolger, 2002; Tein, Sandler, MacKinnon, & Wolchik, 2004). Only one study addressed intracluster correlation (Kam et al., 2004) but this study lacked adequate power to detect program effects.

Third, several studies did not measure effects on theoretical mediators. For example, although three out of the four programs include curricula relating to emotion regulation, only one study evaluated this outcome (Greenberg et al., 1995). Moreover, although the *ACP* sought to promote means-end thinking, this skill was not measured in any evaluations of this program. Alas, several studies did not measure social cognitive skills at all (Kam et al., 2004; Greenberg et al., 1995; Grossman et al., 1997; Lochman & Curry, 1986; Lochman & Lampron, 1988; McMahon et al., 2000, McMahon & Washburn, 2003; Taub, 2002). Without employing theory-based evaluation designs, intervention researchers lack the ability to determine critical program ingredients (Birckmayer & Weiss, 2000).

Finally, few studies tested mediating and/or moderating effects. Out of the 16 studies reviewed, only two studies (Hudley & Graham, 1993; McMahon & Washburn, 2003) tested the mediating effects of socioemotional skills on posttest aggression and only three studies assessed the impact of moderating variables such as gender or risk status (Greenberg et al., 1995, Orpinas, et al. 1995, Van Schoiack et al., 2002). Given that girls, relative to boys, tend to have higher pretest levels of emotional understanding (Bohnert, Crnic, & Lim, 2003) and social information-processing skill (Coie & Dodge, 1998), it is possible that intervention effects in these areas are too small to produce behavioral change.

Chapter IV

The *Making Choices* Program

The *Making Choices (MC)* Program is a universal school-based intervention that attempts to minimize social-cognitive and emotional antecedents of aggression and strengthen children's skills for positive peer relations. Although initially designed for use with the third grade (Fraser et al., 2000), the curriculum has been adapted for preschool children and pre-adolescents. The program has been implemented by intervention specialists as well as by teachers and has been delivered to small, mixed groups and whole classrooms. So that it may be easily modified by practitioners for use with different populations, the program manual incorporates group tips to help practitioners adapt the curriculum for populations of different ages, cultures, and economic backgrounds. In addition, lesson activities include male- and female-normative examples of aggression.

Intervention Strategies

As mentioned, two versions of the *MC* program were evaluated in this study. The *MC* intervention consists of seven units that include classroom-based activities designed to promote SIP and emotional regulation skills. The first unit was designed to help students recognize and understand their own feelings and the feelings of others, as well as learn how to cope with these feelings. It is followed by six units designed to teach children how to build skills at each stage of social information processing.

MC+ consists of the classroom curriculum and incorporates increased opportunities for parent and teacher involvement. To involve parents, the *Family Nights* program was delivered. Based on the *Strong Families* curriculum (Fraser et al., 2004), the *Family Nights* program addresses parenting issues and seeks to reduce family stress. It offers didactic training during multifamily group sessions, to interrupt family contingencies (e.g., use of harsh punishment, lack of positive reinforcement) that reinforce aggressive behavior. Informed by the literature on aggression (McDonald, Billingham, Conrad, Morgan, & Payton, 1997; Reid & Patterson, 2002), *Family Nights* teaches parents how to use more effective parenting techniques (e.g., the use of clear requests, consistent discipline, and rewards), attempts to decrease family stress and social isolation, and involves parents in activities with their children to reinforce SIP skills.

Session attendance for *Family Nights* was voluntary. To facilitate program participation, child care, transportation, and food were provided, and sessions were held concurrently in English and Spanish. *Family Night* newsletters containing information about recent and upcoming sessions were mailed to all families participating in the *MC+* program after each session.

Activities designed to increase teacher involvement were also added to the *MC+* program. Classrooms implemented the *Good Behavior Game* (Barrish, Saunders, & Wolfe, 1969), as a classroom behavior management strategy. In addition, teachers were provided supplemental lesson activities, to reinforce *Making Choices* skills during general classroom instruction. These activities were accompanied by short homework assignments that students took home to work on with their parents. Research staff provided teachers with material and organizational assistance for implementing these activities on a regular basis.

Implementation Characteristics

The *Making Choices* curriculum was implemented approximately 45 minutes per week, once a week, over the course of the academic year. The 29-lesson curriculum was delivered over 21 to 23 sessions. Dosage was measured as the percentage of *Making Choices* sessions attended (and not by the number of hours of service delivered). The decision to measure dosage in this way was based on treatment fidelity information indicating that workers covered lesson material at varying rates – often due to differences in the amount of time each school allotted to the program.

Service providers also completed attendance sheets at the beginning of each lesson. Students' level of exposure to the intervention varied broadly, ranging from 0% to 100%. Nine students assigned to the intervention conditions had no exposure to the program because English as a Second Language (ESL) lessons coincided with the time *Making Choices* lessons were delivered. However the majority of students (roughly 84%) participated in at least 90% of the sessions, and 44.2% of students (164 of 371) attended all sessions.

The Family Nights program was delivered over five, 1.5 hour sessions. Participation rates program were low. On the basis of attendance rosters, 28% (55 of 198) of children in MC+ had parents who participated in at least one family night session.

Prior Research

Results from four pilot studies suggest that *Making Choices* is effective in strengthening promotive factors associated with peer acceptance and reducing aggression (Fraser, Day, Galinsky, Hodges, & Smokowski, 2004a; Fraser, Galinsky, Smokowski, Day, Terzian, Rose, & Guo, 2004b; Nash, Fraser, Galinsky, & Cooper, 2003; Smokowski, Fraser, Day, Galinsky, & Bacallao, 2004). The first pilot study tested the first three units of *Making*

Choices in a middle school in central North Carolina (Nash et al., 2003). As a part of routine school administration, the sixth-grade cohort was divided into two "schools within schools," with one-half of students (n=70) receiving *Making Choices* and the other half receiving instruction as usual (n=95). The sample was predominantly female (59%) and European American (69%), and a large proportion (47%) was academically gifted. To estimate program effects, paired-sample *t* tests and hierarchical linear models (HLMs) were used. This study detected effects on encoding and goal clarification for the overall sample, however, no significant effects on *SLA* skills were found for aggressive-rejected and non-aggressive rejected students. The weak impact on behavioral improvement was attributed to three factors: a) variation in the implementation of the program; b) teachers delivered only one-half of the curriculum; and c) negative peer-group influences. Another reason for weak effects may have been the low statistical power of the study. Effects were estimated with multilevel models despite the fact that the Level 2 equation contained only 5 subjects (i.e., the number of homerooms).

The second study implemented the *Making Choices* plus *Strong Families (MC+SF)* program with an indicated sample of children ages 6 to 12 (Fraser et al., 2004). *Strong Families* is an in-home family intervention that targets issues related to parenting, problem solving, and social support. From nine different sites, 115 children (rated by teachers as aggressive and disliked) and their families were recruited, however, due to dropout and missing data, only 75% (n=86) were retained in the analytic sample. Students were randomized to *MC+SF* (n=41) or a wait-list control group (n=45). Participants received services through the after-school programs of YMCAs, Boys and Girls Clubs, and churches. Program effects were estimated using MANOVA. Findings revealed *MC+SF* to result in

significant improvements in prosocial behavior, emotion regulation, social contact, cognitive concentration, and relational aggression.

The third study was conducted with a universal sample of third-graders (N=101), using a randomized, control group design. Four third-grade classrooms from one rural school in the Southeast were randomly assigned to either *MC* (n=51) or a routine-services control condition (n=50). The ethnic/racial composition of the sample was 68% European American, 22% African American, 5% Latino, and 3% Native American or Asian. Approximately one-quarter of the school population qualified for free-and-reduced lunch. Compared to control group students, *MC* students had higher posttest levels of social contact, social competence , cognitive concentration, and peer acceptance (Smokowski et al., 2004). This study also detected a significant interaction between program assignment and pretest risk status, with higher risk students experiencing greater pretest-to-posttest decreases in overt aggression than lower risk students. Program effects did not vary by gender or race/ethnicity.

Another study was conducted with a universal sample of third graders (N=548) from two rural Southeastern schools (Fraser et al., 2005). The study employed non-randomized cohort design, with three successive cohorts of third graders. The first cohort (2000-01) was exposed to a routine health curriculum, the second cohort (2001-02) received Making Choices (*MC*), and the third cohort (2002-03) received the augmented version of Making Choices (*MC*+). Because students were nested in classrooms (j=29), hierarchical linear models were used to estimate program effects.

Both interventions were found to be effective at mitigating aggression and promoting social competence. Children in *MC* and *MC*+ received lower posttest scores on social aggression, overt aggression, and social competence than children in the comparison

condition. The study detected an unexpected interaction effect for gender, with boys demonstrating greater decreases in overt aggression than girls, controlling pretest levels of aggression and race/ethnicity (Fraser et al., 2005).

In addition, *MC* students were rated significantly higher on social contact at posttest than comparison students, and *MC*+ students received higher posttest ratings of cognitive concentration. A posttest-only study of SIP skills suggested that intervention students were more skillful than comparison students in encoding social cues and setting prosocial goals. In addition, children in the *M*C+ cohort had better scores on hostile attribution and response selection than children in the comparison cohort. A slightly broader pattern of effects was found for *MC*+, suggesting that increasing parent and teacher involvement in structured activities to minimize risk and promote protection is an effective strategy for promoting positive outcomes in elementary-school youth.

The current dissertation research seeks to expand on the most recent study, by utilizing a new comparison cohort to test the main effects of *MC* and *MC*+ on posttest outcomes and testing mediating and moderating effects with this alternate sample. In addition, it conducts a more rigorous analysis of program effects on SIP skills; rather than implementing a posttest-only design, it uses a regressor-variable approach to control for pretest. The following chapter will discuss the methods applied in this study.

Chapter V

Research Methods

The study utilizes a non-randomized, pretest-posttest cohort design with lagged treatment withdrawal. The study sample consists of three cohorts of third graders from two rural elementary schools located in the Southeast region of the U.S. (N=480; see *Table 2*). Cohort 1 (2001-02; n=156) received the *MC* program, and Cohort 2 (2002-03; n=193) received the *MC* program plus supplemental activities requiring greater teacher and parent involvement *(MC+)*. After a one-year treatment withdrawal period, data were collected from a third cohort (2004-05; n=131), who participated in a routine health curriculum.

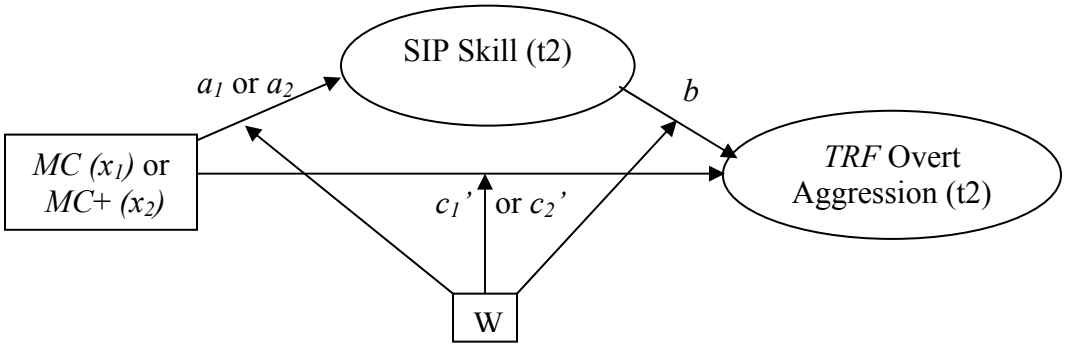
Research Hypotheses

This dissertation research was guided by a number of hypotheses. Compared to comparison students, intervention students were expected to have lower scores on posttest overt aggression and hostile attribution and higher scores on posttest emotion regulation, encoding, goal clarification, and response selection. Gender was expected to moderate program effects on these six outcomes. Posttest SIP skills were expected to partially mediate program effects on teacher-rated posttest overt aggression. *Figure 3* (shown below) summarizes the set of relationships hypothesized by the study:

- main effects of *MC* and *MC+* on all theoretical mediators (a_1 and a_2) and on overt aggression (*c1* and *c2*);

- moderating effects of gender (W) on effects of *MC* and *MC+* on overt aggression (*c1* and *c2* paths) and on effects of SIP skills on aggression (*b* path); and
- mediating effects of SIP skills (partially explaining effects of *MC* and *MC+* on overt aggression; *a1*b* and *a2*b*)

Figure 3: Hypothesized Mediating and Moderating Effects



For ease of illustration, six covariates (male, African American, Latino, pretest overt aggression, pretest SIP skill, and program-by-gender interaction terms) are not shown in this figure. Theoretical mediators were hypothesized to partially mediate program effects on overt aggression. Consistent with findings from the Fraser et al. (2005) study, gender was hypothesized to moderate program effects on overt aggression, with girls displaying less behavioral improvement than boys, on average. Gender was also expected to moderate program effects on SIP skills in the same way, given evidence of gender differences in social cognition (Hughes, et al., 2004). Finally, preliminary research evidence suggests gender differences in the association between SIP skills and aggression (Musher-Eizenman et al., 2004; Schultz et al., 2004).

Sample and Population

Children, parents, and teachers in two elementary schools, located in one rural county in central North Carolina, participated in the study. The study population consisted of three

cohorts of third graders (N=570) from 28 classrooms. Retained students (n=29), however, were excluded from the sample, to ensure that children had not been exposed to more than one condition. From the remaining 541 students, 61 students were excluded due to missing data at pretest and/or posttest on study variables (i.e., 23 from the *CC*, 32 from *MC*, and 11 from *MC*+). Thus, data from 480 students were included in the analysis (Comparison=131; *MC*=156; and *MC*+=193).

Table 2: Sociodemographic Characteristics of the Analytic Sample

		Race/Ethnicity				Sex	
Cohort	n	Latino row % (n)	White row % (n)	Black row % (n)	Other row % (n)	Male row % (n)	Female row % (n)
MC	156	49% (62)	34% (53)	21 % (33)	5% (8)	51% (80)	49% (76)
MC+	193	41% (79)	40% (77)	16% (30)	4% (7)	51% (99)	49% (94)
CC	131	57% (74)	25% (33)	15% (20)	3% (4)	45% (59)	55% (72)
Total	480	45% (215)	34%(163)	17% (83)	4% (19)	50% (238)	50% (242)

The analytic sample is gender-balanced and ethnically-diverse (see *Table 2*). The mean age and ethnic/racial composition was equivalent to the population from which it was drawn (N=546; mean age= 8.7, 46% Latino, 32% European American, 18% African American, and 4% 'Other'). Based on school administrative data, the average percentage of free- and reduced- lunch for the total population represented by both schools was approximately 53% (76.3% for School 1 and 24% for School 2), when averaged across the three cohorts. At both schools, the proportion of Latino students increased with each subsequent cohort, due to increases in the Latino immigrant population in the local community. The age of the analytic sample ranged from 7.2 to 11.6 years old, with a mean of 8.7 years (SD=.66).

Human Subjects Considerations

Whenever children are utilized as subjects of a research study, special ethical and legal considerations must be made. Based on the determination of risk and benefit associated with study participation, informed consent is often needed from parents. In relation to this study, risk and benefits were determined to be minimal by the school administration and by the university-based Institutional Review Board (IRB). Active parental consent was not required because the intervention was officially adopted by the school district and considered to be part of the regular curriculum. All survey measures were seen as part of routine classroom assessment. Moreover, parents were not involved in data collection activities. A letter of support from the school administration was obtained by the Project Coordinator at the commencement of the study. All consent procedures used by the project were reviewed and approved by the school district and the IRB office at the University of North Carolina at Chapel Hill's Office of Academic Affairs.

Data Collection

Pretest/posttest data from the two intervention cohorts (*MC:* n = 185; *MC+*: n = 202) had previously been collected and analyzed (Fraser et al., 2005). Primary data collection was conducted with a new cohort of third graders (n=154); this cohort served as the comparison group for the current study (see *Table 3*).

Table 3: Study Design

Cohorts		2001-02		2002-03		2004-05	
N=541	n	Fall	Spring	Fall	Spring	Fall	Spring
MC	185	T1	T2				
MC+	202			T1	T2		
Comparison	154					T1	T2

Prior to collecting these data, the researcher checked with principals, teachers, and school social workers, to verify that this cohort was not involved with any other form of social skills intervention. Consistent with the previous two cohorts, data were collected during the months of October and April. Data from the original comparison group cohort (2000-01) were not included in this study, due to missing data on pretest SIP skills. No data were collected during the 2003-04 year.

Statistical Power Analysis

The *Optimal Design (OD)* program (Cohen, 1988; Raudenbush & Liu, 2001; Raudenbush, Spybrook, Liu, & Congdon, 2005) was used to estimate power to detect main effects. This software program is specially designed for the estimation of power for multilevel models. Specifically, it is able to determine how many clusters are needed to detect an effect of a given size at a pre-specified α and power level. Power is defined as the ability of a study to reject a null hypothesis of no association between the dependent variable and independent variables (Shadish, Cook, & Campbell, 2002).

To conduct the power analysis, the *cluster-randomized trial* option[1] was selected from the menu of the OD program. Under the heading, *power for main effects of treatment on continuous outcome*, the *power v. # of clusters* option was selected. This option requires the following input: the intraclass correlation (ICC) statistic (ρ), the average cluster size (n), a standardized effect size (δ), and an alpha-level. ICC levels were set to .01, .05, and .15, based on previous research conducted by Carvajal and his colleagues in a paper published in *Multivariate Behavioral Research* (Carvajal, Baumler, Harrist, & Parcel, 2001). On average, the number of students per classroom was 17 (the cluster size). Power was estimated for

[1] This option is required for performing power analyses with random-effects models that evaluate group-level interventions.

small, medium, and large effects. Using the metric for the Cohen's *d* statistic (Cohen, 1988), these effect sizes were set to .20 (small), .50 (medium), and .80 (large).

A power analysis was conducted for an alpha (α) level of .05. Although Bloom (2005) states that the use of one-tailed tests (α=.10) is appropriate when evaluating cluster-randomized social experiments (because they are hypothesized to benefit participants), two-tailed tests are generally preferred. Two-tailed tests of significance are advantageous in that they allow the detection of both positive and negative effects; this is especially important when conducting program evaluation research (Glasgow, Vogt, & Boles, 1999). Findings of the multilevel power analysis are presented in *Table 4*.

Table 4: Power Analysis Information

	Power at α =.05		
Program Effects	*ICC=.01*	*ICC=.05*	*ICC=.15*
Small Effect (.20)	50%	35%	21%
Medium Effect (.50)	100%	98%	82%
Large Effect (.80)	100%	100%	100%

The results of this power analysis suggest that, the study will have adequate power to detect medium and large main effects at all ICC levels (power decreases as ICC increases). However, there is insufficient power to detect small effects, with power estimated at 21%, 35%, and 50%, for ICCs of .15, .05, and .01, respectively; all of which are below the desired threshold of 80% (Cohen, 1988). School-based intervention research studies, which often involve fewer than 50 classrooms, are often affected by the problem of low power to detect small effects.

Instrumentation

Children in all third-grade classrooms were rated by teachers in the Fall (pretest) and Spring (posttest), on a set of six behavioral outcomes measured by the *Carolina Child Checklist – Teacher Form (CCC-TF;* see Appendix C*).* Two *CCC-TF* subscales – emotion regulation and overt aggression – were used in the analysis. Encoding, hostile attribution, goal clarification, and response selection were measured using a child-report measure called the *Skill Level Activity (SLA*; see Appendix D). All scales, and their corresponding items, are described in *Table 5*.

To assess the dimensionality, reliability, and validity of the scales with the current sample, psychometric analyses were conducted (See *Table 6* for results). The dimensionality of each scale was evaluated using single-level Confirmatory Factor Analysis (CFA) in *Mplus 4.1* (Muthén & Muthén, 1998-2006). Because the emotion regulation and overt aggression scales had been validated in prior studies, (Fast Track, 1997; Fraser et al., 2005; Macgowan et al., 2001), a confirmatory approach was deemed appropriate. First-order CFAs were estimated for each scale; a second-order CFA was not conducted due to identification issues. Measures of model fit, such as root mean square errors of approximation (RMSEAs), comparative fit indices (CFIs), and factor loadings were noted.

Briefly defined, the CFI is a "sample-size adjusted measure of fit derived from the comparison of the hypothesized model to the independence model" and the RMSEA is a measure of the "closeness of fit" (Seipel & Apigian, 2005). Factor loadings indicate the extent of factor variance explained by items. CFI values higher than .90 (Browne & Cudeck, 1993; Kline, 1998) and RMSEA values lower than .10 (Browne & Cudeck, 1993) appear to be indicative of adequate model fit, although others have recommended stricter criteria for

model fit (e.g., RMSEA≤.05, and CFI ≥.95; Hu & Bentler, 1999). Measures of dimensionality for each scale are summarized in *Table 6*.

Table 5: Description of Outcome Measures

Variable (# of Items)	Variable Type	Data Collection Instrument	Items
Encoding (6 items)	• Mediating Variable (MV) • Interval level • Average of 6 vignettes with a dichotomous response scale	Items from the Skill Level Activity (*SLA*) Scale, adapted from Home Interview for Attribution Bias; child-rated.	• Look at the picture and circle all of the clues that tell you what is happening.
Hostile Attribution (6 items)	*Same as above*	*Same as above*	• Why did the person in the story do what she or he did? Draw an X on the face you choose.
Goal Clarification (6 items)	*Same as above*	*Same as above*	• If you were the person in the story, what would you want to happen?
Response Selection (6 items)	*Same as above*	*Same as above*	• What would you do?
Emotion Regulation (4 items)	• Mediating Variable (MV) • Interval level • Average of four items rated on a 5-pt Likert scale	Items from the Emotion Regulation subscale of the Teacher Observation of Child Adaptation-Revised (*TOCA-R*); teacher-rated.	• Can calm down when excited or all wound up • Controls temper when there is a disagreement • Expresses needs and feelings appropriately • Very good at understanding other people's feelings
Overt Aggression (5 items)	• Dependent Variable (DV) • Interval level • Average of 6 ordinal items on a 0 to 2 scale	Items originating from Aggression Scale of the Child Behavior Checklist – Teacher Report Form (*TRF*); teacher-rated.	• Bragging, boasting • Cruelty, bullying, or meanness to animals, • Physically attacks people • Teases people • Threatens people

Next, two types of reliability were tested: internal consistency and split-half reliability. A reliability analysis was conducted in *SPSS 14.0* to estimate the Cronbach's Alpha coefficient for each scale. This internal consistency coefficient (α) should be above .90 (Bland & Altman, 1997); however, Nunnaly (1978) has indicated that an alpha-level of .70 is acceptable. The split-half reliability of each scale was estimated using a three-step process. First, each scale was split into two subscales containing an equal number of items; the first subscale was a composite average of the first 2 or 3 items, and the second subscale was a composite average of the last 2 or 3 items. Subscales were constructed separately for pretest and posttest. Second, bivariate correlations between the two subscales corresponding to each scale were run to estimate Pearson's correlation coefficients. Third, correlation coefficients were used to estimate Spearman-Brown Prophecy ($sbp=2r/(1+r)$) statistics for each scale.

Finally, concurrent validity was examined. This type of validity tests whether a measure has "correspondence to a criterion that is known concurrently" (Rubin & Babbie, 2001). To determine concurrent validity, all measures were correlated with the *TOCA-R* Cognitive Concentration scale. This scale was selected due to its empirical and theoretical association with constructs measured in the study (Landau & Moore, 1991; Maedgen & Carlson, 2000). Moderate correlations (r = .40 to r = .70) were expected. Finally, to evaluate construct validity, the convergent validity of the *TRF* Overt Aggression scale was evaluated. To establish adequate convergent validity, two different methods of measuring the same construct must yield similar results (Rubin & Babbie, 2001). Therefore, the correlation between the *TRF* and *TOCA-R* overt aggression scales was examined. A large correlation (higher than .80) between these two measures was expected.

Table 6: Dimensionality, Reliability, and Validity of Measures

Variable	Item-level response scale	Dimensionality			Reliability		Concurrent Validity[2]
		RMSEA	CFI	Standardized Factor Loadings[3]	α Pre/Post	*sbp*[4] Pre/Post	Pre/Post
Encoding	Interval (%)	.06	.98	1: .60, -- 3: .71, $p<.05$ 5: .65, $p<.05$	.71/.66	--	.05/.08+
Hostile Attribut-ion	Nominal (0 or 1)	.00	1.00	1: .86, -- 4: .93, $p<.05$ 5: .72, $p<.05$	.49/.49	--	.05/-.03
Goal Clarific-ation	Nominal (0 or 1)	.00	1.00	2: .91, -- 3: .91, $p<.05$ 4: .81, $p<.05$ 6: .84, $p<.05$	.73/.79	68***/ .69***	.15**/ .16***
Response Selection	Nominal (0 or 1)	.09	.97	1: .71, -- 2: .65, $p<.05$ 3: .81, $p<.05$ 4: .81, $p<.05$ 5: .81, $p<.05$ 6: .80, $p<.05$	.75/.82	.69***/ .81***	.17***/ .17***
Emotion Regulat-ion	Ordinal	.29	.96	1: .74, -- 2: .72, $p<.05$ 3: .81, $p<.05$ 4: .81, $p<.05$	.83/.84	.77***/ .78***	.69***/ .74***
Overt Aggress-ion	Ordinal	.00	1.00	1: .84, -- 2: .93, $p<.05$ 3: .77, $p<.05$ 4: .92, $p<.05$ 5: .85, $p<.05$	.79/.81	.72***/ .74***	-.26***/ -.31***

*$p<$ .05. **$p<$.01. ***$p<$ .001.

[2] To assess concurrent validity, each scale was correlated with the *TOCA-R* Cognitive Concentration scale.

[3] No significance test was conducted on the first item of every scale, because factor loadings were fixed to one.

[4] *sbp* is the abbreviation for the Spearman Brown Prophecy statistic.

Table 6 offers fit indices, factor loadings, internal consistency statistics, split-half reliability estimates, and concurrent validity correlations for pretest and posttest measures in the current study. Except for hostile attribution, all measures were found to be reliable. Sufficient levels of concurrent validity were found for emotion regulation and overt aggression. Low concurrent validity was found for all SIP measures; however low correlations may have occurred as a result of measurement error associated with correlating scales rated by different sources (i.e., child and teacher). More specific information on dimensionality, reliability, and validity is provided in the following paragraphs.

SIP Skills

An adaptation of the Home Interview for Attributional Bias (HIAB; Dodge, 1980), the Skill Level Activity (*SLA*) was used to assess four out of five SIP skills. The HIAB is a group-administered instrument that is designed to assess how children respond to common social situations. More specifically, it measures children's mastery of four SIP skills taught in the Making Choices classroom curriculum (e.g., encoding, interpretation, goal clarification, and response selection). The instrument was administered to each classroom by a member of the research staff who read a series of six short stories describing an ambiguous peer provocation. The stories involve provocations that are relational, e.g., being excluded from a birthday party, physical, e.g., being pushed from behind, or instrumental, e.g., a friend dropping a magazine in the mud (See Appendix D). Each story was accompanied by an illustration of the vignette. Students were asked to look at the picture corresponding to the story and imagine that they were the main character in the story. They were then told that, in each story, "another child does or says something that may affect [the main character] in a good or bad way'. After each vignette, children were asked four questions, each

corresponding to a specific SIP skill (see *Table 5*). They were first asked why they thought the child acted this way, by marking a face marked *friendly*, *mean*, *mistake*, or *can't tell* (interpretation). Second, they were asked to circle all the clues in the picture that told them about what was happening in the story (encoding). Third, they were asked what they "would want to happen" (goal setting), and then, fourth, they were asked "what they would do" in that situation (response selection). Aside from the encoding question, all questions had multiple choice answers. Students were asked to circle the response that best matched how they would respond in that situation. All answers were coded on a dichotomous response scale. Non-aggressive and neutral item responses received a score of 1 and aggressive responses receive a score of 0. Variables were scored by taking the average of all scores for that item across the six vignettes. The hostile attribution scale was reverse-coded so that higher values indicated more aggressive responses.

A technical report of this measure (*Making Choices* Project, 2004) found moderate inter-item correlation (α=.71) for the entire instrument. The inter-rater reliability for the encoding measure is associated with Cohen's Kappas (κ) ranging from .96 to .98 (*Making Choices* Project, 2004), indicating the measure is reliable when scored by multiple raters. Reliability estimates for three out of four SIP scales (i.e., encoding, goal clarification, and response selection) were good, however the internal consistency for the hostile attribution scale was low.

The original measures of encoding, goal clarification, and response selection had fair reliability in the current study, with alpha levels over .72 and *sbp* estimates over .69. Consistent with the prior study, this analysis found a low internal consistency for hostile attribution (α_{pre}= .49 and α_{post}= .34) and low split-half reliabilities (sbp_{pre}= .53; sbp_{post} = .45).

The six-item CFA for hostile attribution did not converge. To reduce measurement error and improve the overall dimensionality of measures, items with factor loadings below .60 were dropped and new scales were constructed. A three-item measure for hostile attribution was constructed using items 1, 4, and 5. Three items were removed from the encoding measure (items 2, 4, and 6) and two items were removed from the goal clarification measure (items 1 and 5). Split-half reliability coefficients could not be estimated for encoding and hostile attribution due to the small number of items.

The first-order model for three SIP variables (encoding, goal clarification, and response selection) appeared to fit the data well (CFIs > .97), indicating good dimensionality. Several RMSEAs were higher than the desired .05 cut-off, but all were below .10. Finally, the concurrent validity between the SIP scales and cognitive concentration was low – ranging from -.02 to .17. This may be due to a lack of agreement between child and teacher reports or due to measurement error in the *SLA* scales. Further assessment of the *SLA* as a measure of SIP skills should be done in order to establish the validity and reliability of the instrument. For example, comparing teacher- and parent- ratings of children's problem solving skill to child-ratings would help to establish the construct validity of the scale.

Emotion Regulation

Children's ability to understand and manage emotions was measured using the *Carolina Child Checklist-Teacher Form (CCC-TF)* (Macgowan, Nash, & Fraser, 2002; see Appendix C for a copy of the instrument). The *CCC-TF* is a 42-item instrument that was designed for use with teachers and developed to measure social and behavioral factors related to aggressive behavior in six- to twelve-year-old children (Macgowan et al., 2002). The

response scale for the measure was a 6-item Likert-type response scale ranging from 0 to 5 (0=*almost never*, 1=*rarely,* 2=*sometimes*, 3=*often*, 4=*very often*, 5=*almost always*). Negative scales (e.g., relational aggression) are reverse-coded to reflect positive social and behavioral functioning.

The *CCC-TF* is an elaboration of the 37-item *Social Health Profile (SHP*, Fast Track, 1997 which is based on the 26-item *Teacher Observation of Classroom Adaptation-Revised* (*TOCA-R*, Werthamer-Larsson, Kellam, & Wheeler, 1991). The *TOCA-R* has been found to have good reliability and validity (for more information, refer the following papers: Fast Track Project, 1997; Kellam, Branch, Agrawal, & Ensminger, 1975; Werthamer-Larsson et al., 1991). The *CCC-TF* was originally tested with a fairly diverse sample of 171 sixth-grade students (Mean=11.8, *SD*=.38), ages 11 to 13 (Macgowan et al., 2002). The internal consistency of the entire scale was high ($\alpha = .95$) and the test-retest reliability of each scale was moderate (.70-.80). The criterion-related and construct validity of the instrument was good. A test of the concurrent validity found a .64 correlation with grade point average, and a test of convergent validity found a -.77 correlation with pretest scores on the *TRF* (Macgowan et al., 2002).

The dimensionality of the emotion regulation subscale received adequate support (i.e., CFI=.96; significant factor loadings), although the RMSEA value was high (RMSEA=.29). The scale had good reliability and validity. Cronbach Alpha's for pretest and posttest were .83 and .84, respectively. Spearman Brown Prophecy (*sbp*) estimates were also good (sbp_{pre}=.77 and sbp_{pre}=.78). Finally, the scale had an adequate level of concurrent validity (r_{pre}= -.69; r_{post}= .74).

Overt Aggression

The *CCC-TF* was also used to measure overt aggression. It includes 24 items from the *Teacher Report Form (TRF)*. The *TRF* is a standardized teacher assessment of social competence, adaptive functioning, academic performance, and social and behavioral problems. It was originally normed on a representative sample of 1391 children. Studies have found adequate reliability and validity for the overall measure (Achenbach, McConaughy, & Howell 1987). A high internal consistency coefficient (α=.97) has been reported for children ages 4 to 11 (Greenhill & Malcolm, 2000). Composite behavior scores were found to have good stability, with one-week, test-retest correlations of .97 for competence scores and .92 for problem behavior scores (Greenhill & Malcolm, 2000). Generally, correlations of over .40 across a 3-12 month period are considered adequate (Robinson, Shaver, & Wrightsman, 1991). The *TRF* has also been found to have substantial criterion-related validity, with cross-validation techniques yielding significant correlations for each subscale.

Overt aggression was measured using a 6-item narrow-band subscale, derived from the aggression subscale of the externalizing problem behavior scale of the *TRF* (Achenbach & Edelbrock, 1991; Fast Track Project, 2003). This subscale was used rather than the 24-item scale, in order to obtain a more conceptually-distinct construct for overt aggression that excludes items related to oppositional-defiant behavior (e.g., "Argues a lot" and "Defiant, talks back to staff") and hyperactive/disruptive behavior (e.g., "Talks too much" and "Unusually loud"). The scale is a composite average of six ordinal response items (see *Table 5*). Teachers rate student behavior based how true the item reflects student's behavior during the last month (0=not true, 1=somewhat or sometimes true, 2=very true or often true). The

Cronbach's Alpha for this subscale was .81 in a sample of kindergarten children (Fast Track Project, 2003).

In the current study, the factor-structure of a five-item narrow-band subscale was supported (item 12, "fights with others," was dropped due to a negative between-level variance found after performing a multilevel CFA – for more detail, see page 110). All items loaded significantly and in the expected direction onto one factor. In addition, measures of fit were acceptable (CFI=.99; RMSEA=.06). The internal consistency measure was acceptable at .81(pre) and .79(post). The split-half reliability coefficients were also adequate (sbp_{pre} = .72; sbp_{post} = .74). Sufficient levels of concurrent validity (r_{pre} = -.26; r_{post}= -.31) were obtained. Finally, strong support was obtained for the convergent validity of this measure. A strong correlation was obtained when correlating the *TRF* Overt Aggression measure with the *TOCA-R* Overt Aggression subscale (r_{pre} = .67, p<.001; r_{post} = .68, p<.001).

Data Analysis

Multilevel structural equation modeling (MSEM) was used to estimate main, moderating, and mediating effects. Multilevel models are generally recommended when working with grouped data and larger samples, especially when assessing the impact of group-level predictors such as program assignment (Krull & MacKinnon, 1999; Nash, Kupper, & Fraser, 2004). Conducting a single-level analysis in this context is inappropriate because it is likely to result in downwardly biased standard errors and increase the probability of making a Type I error (Krull & MacKinnon, 2001, pp. 264). In addition, ignoring the effects of clustering violates the assumption of uncorrelated or independent error.

Prior to conducting the multilevel explanatory analyses associated with the primary research questions of this study, a series of descriptive analyses were performed. These analyses were conducted to evaluate potential threats to internal validity. First, the amount of unexplained variation due to clustering was estimated. Clustering effects were measured by estimating intraclass correlation (ICC) coefficients which measure the proportion of total unexplained variance explained by between-group variability (Killip, Mahfoud, & Pearce, 2004). Second, an attrition analysis was conducted to assess differences between students participating in the study (the target sample) and students included in the final study sample (the analytic sample). When a sample does not adequately reflect the population from which it is drawn, this is referred to as *sampling bias* (Johnson, Beaton, & Murphy, 2005). This analysis, however, examined differences between the sample and the subsample, so inferences about the general population of third-grade children in the U.S. cannot be drawn.

Third, due to the nonrandomized study design, sources of *selection bias* were examined by estimating between-group differences on observed covariates. Selection bias occurs when there are differences between intervention conditions on measured and/or unmeasured factors that may also be related to differences in a dependent variable or "outcome differences" (Shadish, Cook, & Campbell, 2002). Selection bias violates the OLS regression assumption of non-recursivity, because the treatment effect is correlated with unexplained variance in the dependent variable; this is violation also referred to as *endogeneity bias* (Ettner, 2006). Selection effects result in biased estimates of treatment effects. For instance, if a group assigned to receive social skills training has lower social competence scores than a control group at pretest, treatment effects are likely to be inflated.

Estimating Intraclass Correlations

Intraclass correlation (ICC) coefficients were estimated for each dependent variable, in order to assess for clustering effects. In a two-level model, this statistic (ρ) is typically estimated by dividing the level 2 variance component by the sum of the level 1 and level 2 variance components (i.e., $\rho = \tau^2 /(\sigma^2 + \tau^2)$) . In a two-level model, the level 1 (residual) variance component represents the within-group variability and the level 2 (group-level) variance component represents the between-group variability. The variance components (σ^2 and τ^2) can be found in the covariance parameter estimates output of *SAS Proc Mixed* (SAS, 2000), or in the *Mplus* output under the headings: within-level variance and between-level variance. ICCs are automatically calculated in *Mplus,* but they must be hand-calculated when using *SAS Proc Mixed*.

Table 7: Intraclass Correlations

Variable of Interest	ICC for Unconditional Random Intercept Model[5]	ICC for Conditional Random Intercept Model
T2 Overt Aggression	.09	.12
T2 Emotion Regulation	.23	.21
T2 Encoding	.09	.09
T2 Hostile Attribution	.02	.00
T2 Goal Clarification	.05	.01
T2 Response Selection	.04	.00

Note. Unconditional models were regressed on pretest only.

Conditional and unconditional random intercept models with the analytic sample were estimated, to obtain two different ICC estimates. As in the final models estimated in the explanatory analysis, the conditional models included theoretical covariates such as Male,

[5]Unconditional models were conditioned on pretest only.

Latino, African American, and pretest[6], as well as dichotomous intervention indicators for *MC* and *MC+*. The models for overt aggression and goal clarification differed from the final models in that they excluded interaction terms. The ICCs resulting from the conditional and unconditional models are reported in *Table 7*.

Traditionally, ICC levels of .05, .10, and .15 are considered small, medium, and large (Hox, 2002). According to Krull and MacKinnon (2001, p. 261), multilevel analysis is warranted when the ICC is greater than or equal to .15. Other scholars have argued for lower cut-off points, stating that, whenever interventions are delivered at the group level, there is some amount of group-level variation that must be accounted for (Bloom, 2005). In their evaluation of the *Safer Choices* study, Carvajal et al. (2001) demonstrated that ICC's as low as .01 produced a biased test of program effects. Zyzanski, Flocke, & Dickinson (2004) argue that small ICC levels of .05 are high enough to warrant multilevel methods: "small intracluster correlations coupled with large cluster size can still affect the validity of conventional statistical analyses" (Zyzanski et al., 2004, p. 200).

One guideline for deciding whether ICC is high enough to affect parameter estimates is if the design effect is over 2.0 (Muthén, 1999). The design effect (DEFF) is the ratio of the actual variance to the variance computed under the assumption of simple random sampling, and is estimated using the formula: $DEFF = 1 + \delta (n - 1)$, where n is the cluster size and δ (rho) is the intraclass correlation (Shackman, 2001). In this study, an ICC of .06 or more was found to produce a design effect. This cut-off was used as the critical value for determining the need for a multilevel analytic approach. As shown in *Table 7*, the ICCs for overt aggression, emotion regulation, and encoding were high enough to warrant the use of

[6]Pretest variables corresponded to whichever variable was being tested (e.g., pretest emotion regulation was included as a covariate in the model for posttest emotion regulation).

multilevel models. The ICCs for goal clarification, response selection, and hostile attribution were below the cut-off point. However, goal clarification was modeled using a multilevel approach because there was sufficient between-level variance to allow the estimation of the multilevel model. Thus, significance tests for effects on goal clarification may be slightly conservative.

Attrition Analysis

Attrition rates for each cohort and time point were estimated (see *Table 8*). Two attrition analyses were performed to compare the characteristics of intervention participants with teacher- and child-rated data at both time points (n=480) to the characteristics of intervention participants with missing data at any time point (n=61). For simplification purposes, cases with missing data will be referred to as *attrited* and cases without missing data will be referred to as *non-attrited*.

The first analysis compared attrited and nonattrited participants on pretest behavioral and social-cognitive measures using independent-samples *t*-tests (see *Table 9*). The Levene's Test was used to test inequality of variances between means. If this test was significant ($p<.05$), the *p*-value of the unequal variances test statistic was reported. Results suggested

Table 8: Data Attrition by Time and Cohort

	Cohort 1 *(MC)*		Cohort 2 *(MC+)*		Cohort 3 *(CC)*	
	No Missing n (%)	Missing n (%)	No Missing n (%)	Missing n (%)	No Missing n (%)	Missing n (%)
Pretest	173 (93.5%)	12 (6.5%)	199 (98.5%)	3 (1.5%)	139 (90.3%)	15 (9.7%)
Posttest	166 (89.7%)	19 (10.3%)	195 (96.5%)	7 (3.5%)	145 (94.2%)	9 (5.8%)

Table 9: Attrition Analysis of Pretest Measures

	No Missing		Missing		
	Mean (SD)	N	Mean (SD)	N	*t*-statistic
Child Characteristics					
Age	8.68 (.66)	480	8.77 (.75)	61	-.99
Behavioral Scales					
Authority Acceptance† (0-5)	4.41 (.57)	480	4.23 (.74)	52	-1.61
Social Contact (0-5)	3.94 (.67)	480	3.81 (.74)	52	-1.33
Cognitive Concentration (0-5)	3.23 (.98)	480	2.88 (1.09)	52	-2.37
Social Competence (0-5)	3.25 (.83)	480	3.11 (.93)	52	-1.15
Emotion Regulation(0-5)	3.28 (.90)	480	3.20 (1.06)	52	-.65
TRF Overt Aggression† (0-2)	.10 (.23)	480	.16 (.32)	52	1.37
SIP Scales (0-1)					
Hostile Attribution	.59(.35)	480	.54 (.37)	36	-.81
Encoding	.45 (.19)	480	.38 (.23)	36	-2.21*
Goal Clarification	.80 (.31)	480	.83 (.28)	35	.47
Response Selection	.67 (.32)	480	.66 (.31)	35	-.06

Note. A dagger sign (†) is used to indicate scales with unequal variances.

*p< .05. ** p<.01. *** p< .001.

Table 10: Attrition Analysis of Race/Ethnicity and Sex

Variable	Not Missing	Missing	Total	χ2-statistic (df, n)
Race/Ethnicity - % (n)†				
EA	34.2% (158)	22.8% (18)	32.5% (176)	3.96[7]
AA	16.0% (74)	27.8% (22)	17.7% (96)	(2, N=541)
Latino	45.9 % (212)	46.8% (37)	46.0% (249)	
Other	3.9% (19)	2.5% (2)	3.7% (20)	
Sex - % (n)				
Male	48.7% (225)	67.1% (53)	51.4% (278)	6.78**
Female	51.3% (237)	32.9% (26)	48.6% (263)	(1, n=541)

*p< .05. **p<.01. ***p< .001.

that attrited students had lower pretest levels of encoding (t(514)=2.21, p<.05) than non-attrited students. Because these groups did not differ on any other behavioral and social

[7] Between-cell differences for race/ethnicity were tested using a three-group race/ethnicity variable (i.e., White/Black/Other).

cognitive measures, this was not viewed as posing a significant threat to equivalence of cases with and without missing data.

The second analysis examined sociodemographic differences between attrited and nonattrited students, by using a crosstabs analysis, which tested between-cell differences using the Pearson's Chi-Square Test (see *Table 10*). This method tests the alternative hypothesis that row and column variables are independent and is appropriate for comparisons between nominal variables with two or more categories and more than five observations per cell (Stokes, Davis, & Koch, 2000). Chi-square tests detected a significant difference in gender composition, with a higher proportion of males in the attrited group than in the non-attrited group (i.e., 67% versus 50%). No significant differences between attrited and non-attrited students in overall ethnic/racial composition were found.

Selection Bias Analysis

In order to test the equivalence of study conditions, two selection bias analyses were conducted, using cohort assignment as the grouping variable (N=480). Between-cohort differences in pretest means were assessed using a one-way analysis of variance (ANOVA) test in *SPSS 14.0* for Windows (see *Table 11*). Pairwise multiple comparison tests were conducted when the *F* or Welch statistic was significant, in order to test differences between each pair of means (*Table 12*). When equal variances were assumed, the Least Significant Difference (LSD) test was used. The LSD test uses *t*-tests (α=.05) to perform pairwise comparisons between group means and may only be used when there are three groups (Green & Salkind, 2003). When equal variances were not assumed, the Tamhane's T2 test (a more conservative test of pairwise comparisons) was utilized.

Table 11: Selection Bias Analysis of Pretest Measures

Variable (Response Scale)	*CC* Mean (SD)	*MC* Mean (SD)	*MC+* Mean (SD)	Total Mean (SD)	F or Welch Statistic	df
Child Characteristics						
Age	8.02 (.56)	8.93(.54)	8.92 (.49)	8.68 (.66)	141.45***	(2,477)
Behavioral Scales						
Authority Acceptance (0-5)	4.40 (.58)	4.43 (.58)	4.39 (.56)	4.41 (.57)	.18	(2,477)
Social Contact (0-5)	4.07 (.61)	4.01 (.65)	3.81 (.71)	3.94 (.67)	6.92***	(2,477)
Cognitive Concentration (0-5)	3.21 (1.00)	3.29 (.99)	3.19 (.97)	3.23 (.98)	.54	(2,477)
Social Competence† (0-5)	3.34 (.74)	3.23 .83)	3.21 (.88)	3.25 (.83)	1.08	(2, 309.1)
TRF Overt Aggression† (0-2)	.07 (.19)	.13 (.29)	.12 (.28)	.11 (.26)	2.48	(2, 312.1)
SIP Scales						
Hostile Attribution† (0-1)	.58 (.34)	.50 (.37)	.66 (.33)	.59 (.35)	9.19***	(2, 296.6)
Encoding (0-1)	.55 (.17)	.39 (.17)	.44 (.19)	.45 (.19)	27.40***	(2,477)
Goal Clarification (0-1)	.80 (.32)	.79 (.32)	.82 (.30)	.80 (.31)	.38	(2,477)
Response Selection (0-1)	.69 (.31)	.60 (.34)	.71 (.31)	.67 (.32)	6.01**	(2,477)
Emotion Regulation (0-5)	3.36(.84)	3.23 (.85)	3.28 (.96)	3.28 (.90)	.69	(2,477)

Note. A dagger sign (†) is used for scales with unequal variances; for these scales, the Welch test of significance was used.

Table 12: Pairwise Multiple Comparison Tests for Pretest Differences

Variable	CC v. MC (Sig.)	CC v. MC+ (Sig.)	MC v. MC+ (Sig.)
Age	CC<MC (***)	CC<MC+ (***)	MC≈MC+ (*ns*)
Social Contact	CC>MC (***)	CC≈MC+ (*ns*)	MC>MC+ (***)
Hostile Attribution†	CC≈MC (*ns*)	CC>MC+ (trend)	MC>MC+ (***)
Encoding	CC<MC (***)	CC<MC+ (***)	MC>MC+ (*)
Response Selection	CC>MC (*)	CC≈MC+ (*ns*)	MC<MC+ (***)

Note. A dagger sign (†) is used to indicate scales with unequal variances; for these scales, the Tamhane T2 test is performed. *p< .05. **p<.01. ***p< .001.

Table 13: Selection Bias Analysis: Sociodemographic Characteristics

Variable	*CC* % (n)	*MC* % (n)	*MC*+ % (n)	Total	χ2 (df, n)
Race/Ethnicity					
EA	25.2% (33)	34.0% (53)	39.9% (77)	34.0% (163)	13.36*
AA	15.3% (20)	21.2% (33)	15.5% (30)	17.3% (83)	(6, n=480)
Latino	56.5% (74)	39.7% (62)	40.9% (79)	44.8% (215)	
Other	3.1% (4)	5.1% (8)	3.6% (7)	4.0% (19)	
Sex					
Male	45.0% (59)	51.3% (80)	51.3% (99)	49.6% (238)	1.49
Female	55.0% (72)	48.7% (76)	48.7% (94)	50.4% (242)	(2, n=480)

Note. A difference in proportion test revealed significant race/ethnicity differences 1) between *CC* and *MC* conditions and 2) between *CC* and *MC*+ conditions.
*p< .05. **p<.01. ***p< .001.

Significant between-cohort pretest differences on age, race/ethnicity, social contact, hostile attribution, encoding, and response selection were found (see Tables 11, 12, and 13). Members of the *MC*+ cohort had lower levels of social contact at pretest than the other two cohorts (p<.001). In addition, children in the *MC*+ cohort had lower pretest levels of hostile attribution than children in *MC* (p<.001) and in the comparison group (p<.10). Children in the comparison cohort were younger (p<.001) and had less encoding skills (p<.001) than children in the *MC* and *MC*+ conditions. Finally, Children in *MC* selected more aggressive responses than children in the comparison (p<.05) and *MC*+ cohorts (p<.001).

Differences in sociodemographic characteristics by cohort were tested using Pearson's Chi-Square test (see *Table 13*). A significant chi-square test detected between-cohort differences in the ethnic composition of the sample (χ^2(6, n=480)= 13.35, p<.05). A difference in proportion test revealed significant race/ethnicity differences 1) between *CC* and *MC* conditions and 2) between *CC* and *MC*+ conditions. Further examination revealed

that comparison students were more likely than intervention students to be Latino and less likely to be European American (see *Table 13*).

These mixed findings prohibit us from drawing any conclusions about the presence of systematic differences in SIP skills between the comparison and intervention cohorts at pretest. To mitigate potential selection bias, pretest scores were included as covariates in the analytic models. To examine the influence of ethnic differences on pretest scores, an independent samples *t*-test was performed using an indicator variable for Latino (0=non-Latino; 1=Latino) as the grouping variable. This analysis found no significant pretest differences between Latino students and non-Latino students on all measures. Thus, between-cohort differences in the proportion of Latino students were not expected to bias the estimation of treatment effects.

Multilevel Structural Equation Modeling (MSEM) Analysis

Main, mediating, and moderating effects were tested using multilevel structural equation modeling (MSEM) and single-level, general structural equation modeling (GSEM) in *Mplus 4.1*. A multilevel approach to estimating effects was utilized, because a significant ICC for overt aggression was found. The random=twolevel option was used to estimate all multilevel effects. This option allows intercepts to vary between groups and estimates level 1 and level 2 variance components. The following section details the methods used for estimating these models.

Testing Main Effects. Effects for *MC* and *MC+* were estimated for six variables: overt aggression and five theoretical mediators – i.e., encoding, hostile attribution, goal clarification, response selection, and emotion regulation. A regressor-variable approach was used in this analysis (Allison, 1990). This approach, which partials out the effect of pretest

when measuring the impact of a predictor on a posttest outcome variable, was selected due to the need to control possible selection bias between intervention and control groups (Gillespie & Streeter, 1994) and the need to account for error associated with regression toward the mean (Hsu, 1989). Given numerous studies showing that program effects are often moderated by pretest, constraining the effect of pretest to a fixed constant (i.e., one) by using a change score approach did not make sense. *Figure 4* displays the multilevel model used to estimate effects on overt aggression.

Figure 4: Baseline Multilevel Equation to Test Main Effects

Level 1: Y_{ij} (OVAGG2) = β_{0j} + β_{1j}(MALE) + β_{2j}(BLACK) + β_{3j}(LATINO) + β_{4j}(OVAGG1) + β_{c1j}(*MC*) + β_{c2j}(*MC*+) + r_{ij}

Level 2: β_{0j}= γ_{00} + u_{0j}

In addition to controlling for pretest, the baseline model included gender and race covariates (i.e., MALE, BLACK, and LATINO), and two dichotomous program indicators (*MC* and *MC*+). This model specification is consistent with our prior studies.

One feature of this model that differed from the prior study is the estimation of program effects on the student level (Level 1) and not the classroom level (Level 2). Thus, instead of estimating classroom-level effects on overt aggression, the model predicts student-level effects. The decision to define the program variables as Level 1 variables resulted from the desire to be consistent with models estimating indirect effects, which were unable to estimate indirect effects of the program when program variables were defined as between-level variables. To assess whether this modeling approach affected the magnitude or significance of program effects, the author estimated all main-effects models with program variables included in the Level 2 equation.

Testing Moderating Effects. After estimating the reduced models, moderating effects were tested by adding two within-level interaction terms, β_{5j}(*MC**MALE) and β_{6j}(*MC*+*MALE), to the model. Consistent with the recommended strategy for testing fixed effects in HLM (Raudenbush & Bryk, 2002), these interaction effects were tested using a deviance test approach (except when models were saturated). A normal-theory chi-square difference test was used to test interaction effects in the single-level models [(i.e., χ^2_{nested} - χ^2_{full} / ($df_{nested} - df_{full}$)]. For multilevel models estimated using Restricted Maximum Likelihood (REML), moderating effects of gender on the *a* and *b* paths were tested by performing a Satorra-Bentler Scaled Chi-Square Difference Test (Satorra, 2000). This method divides the normal-theory chi-square statistic "by a scaling correction to better approximate chi-square under non-normality" [(i.e., χ^2_{nested} - χ^2_{full} / ($df_{nested} * cd_{nested} - df_{full} * cd_{full}$); Mplus, n.d.]. When program-by-gender interaction effects were significant, main effects were derived from the full model, rather, than the reduced model, and effect sizes for *MC* and *MC*+ were calculated separately for females and males by running the models twice (once with male=1 to obtain program effects for females, and once with male=0, to obtain program effects for males).

Parameter estimates generated from the final model (full or reduced) were used to estimate effect sizes (standardized deltas - δ) for the interventions. Informed by Raudenbush et al. (2004), effect sizes were calculated by dividing the parameter estimates for the fixed effects of *MC* and *MC*+ by the square root of the total (residual- and classroom-level) variance ($\delta = \beta/[(\tau^2 + \sigma^2)^{1/2}]$).

Testing Mediating Effects. This study tested the mediation of SIP-skills on program effects for overt aggression (see *Figure 5*). Indirect effects were estimated using an

unstandardized product of coefficients method (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002) and tested using the Sobel Test (Sobel, 1982). This test of mediation estimates the standard error of the indirect effect as the square root of the asymptotic variance (MacKinnon et al., 2002). The *z*-statistic for this test is calculated by dividing the indirect effect (b_{ab}) by the following formula: $s_{ab} = \sqrt{b^2 s^2_a + b^2 s^2_b}$ (MacKinnon et al., 2002, p. 85).

In their simulation study examining 14 different methods of testing mediation, MacKinnon and his colleagues (2002) found this test to have greater power than the causal steps approach (Baron & Kenny, 1986), but insufficient power for detecting small effects, especially with small sample sizes. To compensate for power issues, they constructed a cumulative frequency distribution for indirect effects estimated using a product of coefficients approach (b_{ab}/s_{ab}). The critical *z* value associated with a sample size of 50, an alpha of .04, a dichotomous independent variable, and a null hypothesis of no mediated effect was .97 (see http://www.public.asu.edu/~davidpm/ripl/freqdist.pdf). This cut-off was used to assess significance in the current study.

Joint effects were estimated using a structural equation modeling approach. A restricted maximum likelihood approach (REML) was used to estimate model parameter estimates in MSEM, and a maximum likelihood approach (ML) was used to estimate parameter estimates in GSEM. A system of two multilevel equations was specified using the *Mplus 4.1* statistical modeling program (Muthén & Muthén, 1998-2006). These equations are shown in *Figure 6*.

Figure 5: Multilevel Structural Equation Model

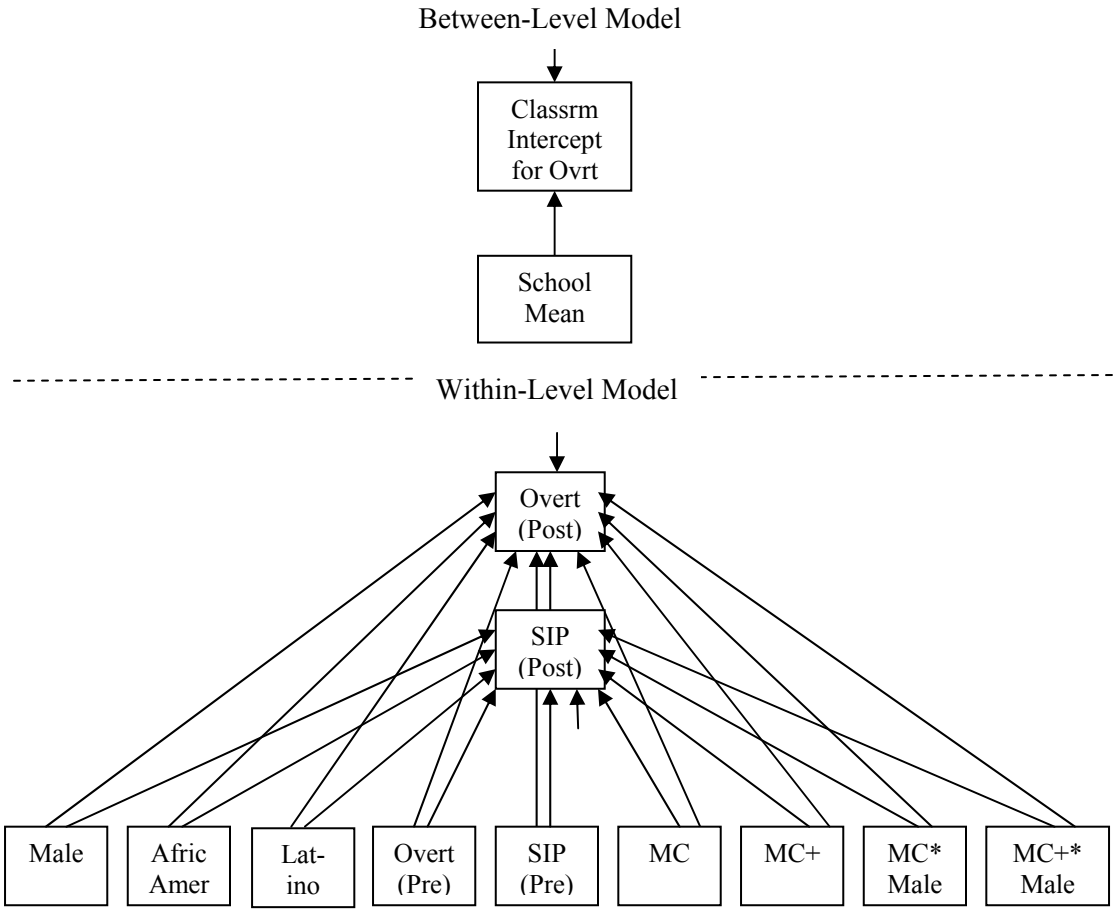


Figure 6: System of Equations to Test Mediation

Equation 1: Y=X +M

Level 1: Y_{ij} ($OVERT_{post}$) = β_{0j} + β_{1j}(MALE) + β_{2j}(AFRIC) + β_{3j}(LATINO) + β_{4j}($OVERT_{pre}$) + β_{5j}(SIP_{pre}) + β_{bj}(SIP_{post}) + $\beta_{c1'j}$(*MC*) + $\beta_{c2'j}$(*MC*+) + β_{6j}(*MC**MALE) + β_{7j}(*MC*+*MALE) + r_{ij}

Level 2: β_{0j}= γ_{00} + u_{0j}

Equation 2: M=X

Level 1: M_{ij}(SIP_{post})= β_{0j} + + β_{1j}(MALE) + β_{2j}(AFRIC) + β_{3j}(LATINO) + β_{4j}($OVERT_{pre}$) + β_{5j}(SIP_{pre}) + $\beta_{a1'j}$(*MC*) + $\beta_{a2'j}$(*MC*+) + β_{6j}(*MC**MALE) + β_{7j}(*MC*+*MALE) + r_{ij}

Level 2: β_{0j} = γ_{00} + u_{0j}

Equation 1 estimates the direct effects of SIP skill (M) and program variables (X_1 and X_2) on overt aggression (b_b, $b_{c1'}$, $b_{c2'}$). Equation 2 estimates the direct effects of *MC* and *MC*+ on SIP skill (b_{a1} and b_{a2}). Each equation represents a two-level model, with Level 1 containing student-level observations, and Level 2 containing observations clustered by third-grade classroom. To facilitate the estimation of indirect effects, the program variables were identified as Level 1 variables. Although this is not ideal, the model to estimate the indirect effects would not run without changing the level of the program from the classroom level to the student level.

Note that the model for overt aggression (Equation 1) includes one additional covariate – the pretest mean for the mediating variable – and one additional predictor – the posttest mean of the mediating variable. Program-by-gender interaction terms in the equation predicting overt aggression account for moderation of the *c1* and *c2* paths. Program-by-gender interaction terms in the equation predicting the theoretical mediator account for moderation of the *a1* and *a2* paths. An additional interaction term was added to Equation 1 to test moderation of the *b* path; however it was later eliminated from the final models due to the fact that it greatly compromised model fit (more detail on this decision is offered in the Results chapter, in the section on moderating effects).

Chapter VI

Results

The study findings reported in this chapter are organized in the order of research questions (i.e., main effects, moderating effects, and mediating effects) listed on page 16. The discussion of main effects summarizes findings related to the effects of *MC* and *MC*+ on posttest overt aggression and posttest social cognition (i.e. SIP-related skills), holding pretest and all other covariates constant. Next, the moderating impact of gender on program effects is discussed. Finally, the discussion of mediating effects reports findings relating to the significance of indirect effects. First, bivariate correlations are presented to provide a context for research findings.

Bivariate Correlations

Prior to estimating the analytic models, bivariate correlations between study variables were examined to assess the magnitude and direction of coefficients (see *Table 14*). Examining the direction of correlation coefficients was done as a construct-validity check. Unusually high correlations were identified as a diagnostic strategy. High correlations between independent variables produce the statistical problem of collinearity, and, consequently, unreliable regression coefficients. Identifying highly correlated variables prior to the analysis can help improve model specification. Weak correlations between independent and dependent variables may relate to errors in data entry, measurement error, or a weak conceptual model.

Nearly all variables correlated with each other in the expected directions, however the majority of correlations were weak. Consistent with the research evidence, hostile attribution was positively associated with overt aggression (r=.13, p<.001), negatively correlated with setting neutral or friendly goals (r=-.18, p<.001), and negatively correlated with selecting non-aggressive responses (r= -.20, p<.001) at pretest. Contrary to the theoretical propositions of the SIP model, hostile attribution was positively correlated with the ability to encode a variety of social cues (r=.10, p<.05). Empirical research suggests that hostile attribution should actually lead to the identification of fewer cues, due to a perceptual bias that recognizes hostile cues before friendly or neutral cues.

Consistent with an emotion-integrated SIP model, which posits relationships between emotion-related constructs and online SIP skill, emotion regulation was negatively correlated with overt aggression (r= -.48, p<.001) and hostile attribution (r= -.10, p<.05) and positively associated with goal clarification (r= .18, p<.001) and response selection (r= .19, p<.001) at pretest. Emotion regulation was not associated with children's ability to encode social clues. Preliminary empirical research appears to support the relationship between emotion regulation and emotion processing skill (Schultz, Izard, & Bear, 2004), which is considered part of the encoding step of the SIP model. Anomalous correlations may, in part, be due to measurement error, or due to chance, but are not significant enough to warrant concern.

Several variables were identified as having moderate-to-high correlations. A moderate correlation between pretest and posttest emotion regulation (r=.61, p<.001) was noted. This correlation did not appear to pose a problem in the mediation model for emotion regulation. Of more significant concern was high correlations between MALE and its

Table 14: Correlation Matrix with Selected Study Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. *MC*	1.00																
2. *MC*+	-.57***	1.00															
3. LATINO	-.07	-.06	1.00														
4. BLACK	.07	-.04	-.41***	1.00													
5. MALE	.02	.03	.00	.08	1.00												
6. ENC1	-.22***	-.07	.10*	-.06	-.09*	1.00											
7. ENC2	-.02	-.04	.08	-.09*	-.22***	.37***	1.00										
8. GOAL1	-.03	.04	.05	-.17***	-.25***	-.04	.10*	1.00									
9. GOAL2	-.04	.18***	-.07	-.12***	-.27***	-.11*	-.07	.39***	1.00								
10. HOST1	.16***	-.17***	-.01**	.06	.01	.10*	.03	-.15***	-.11*	1.00							
11. HOST2	.09*	-.20***	.08	.09	.02	.04	.02	-.07	-.16***	.32***	1.00						
12. RESP1	-.15***	.11**	-.01	-.07	-.24***	-.01	.09*	.74***	.37***	-.20**	-.13*	1.00					
13. RESP2	-.08	.20***	-.07	-.06	-.28***	-.07	-.06	.36***	.75***	-.09	-.19	.44***	1.00				
14. OVRT1	.06	.02	-.12**	.13**	.19***	-.02	-.04	-.15***	-.07	.13***	.07	-.12**	-.06	1.00			
15. OVRT2	-.03	-.08	-.05	.13**	.17***	.06	-.03	-.14***	-.16***	.18***	.07	-.09*	-.14***	.53***	1.00		
16. EMO1	-.07	-.03	.01	-.15***	-.29***	.09	.06	.18***	.12**	-.10*	-.13*	.19***	.14***	-.48***	-.34***	1.00	
17. EMO2	.05	.05	-.05	-.19***	-.26***	.01	.07	.15***	.15***	-.06	-.12**	.14***	.17***	-.29***	-.53***	.61***	1.00

*p< .05. ** p<.01. *** p < .001

interaction with mediating variables measured at posttest (i.e., male*mediating variable). The correlations between MALE and MALE*EMO REG, MALE*ENCODING, MALE*GOAL, MALE*RESPONSE, and MALE*HOS ATT interaction terms were 93, .92, .86, .80, and .75, respectively. These interaction terms were entered as predictors into the mediation models, to testing whether gender moderates the *b* path. However, because adding these interaction terms significantly compromised model fit, a decision was made to exclude these interaction terms from model estimating indirect effects.

Do MC and MC+ Result in Greater SIP Skill and Decreased Overt Aggression?

Single-level and multilevel models were used to estimate main effects of the program on each variable of interest. Single-level models were used for models estimating effects on hostile attribution and response selection, which did not converge when a multilevel approach was used. This occurred due to the fact that these variables lacked sufficient between-level variance. Output generated from running multilevel models for these variables read: "The estimated between covariance matrix is not positive definite as it should be. Computation could not be completed. The variance of [name of DV] approaches zero." An explanation for the lack of group-level variance for hostile attribution and response selection is not apparent, but findings suggest they are less sensitive to contextual effects.

Two-level models were used to estimate main effects on the remaining SIP variables (emotion regulation, encoding, and goal clarification) and overt aggression. Although a three-level (student, classroom, teacher) model of overt aggression fit best in the previous study of posttest findings (Fraser et al., 2005), the two-level (student, classroom) model proved to be a better fit to the data in this study ($BIC_{2\text{-level}}=9.4 < BIC_{3\text{-level}}=10.2$) when random intercept models for overt aggression were estimated using SAS Proc Mixed. This change in random

effects could possibly be attributed to greater consistency between teachers in their behavioral ratings of comparison-group children; it may also be due to differences between study samples. All models included fixed effects for the outcome-specific pretest, male, African American, Latino, *MC, MC+,* and a random effect for the Level 2 intercept. All fixed effects were grand-mean centered, except for component effects of the interaction terms (i.e. male, *MC, MC+*), in order to obtain meaningful program effects for males and females in each intervention condition. The final models for goal clarification and overt aggression included significant program-by-gender interaction terms (i.e., Male**MC* and Male**MC+*). *Table 15* depicts the fixed and random effects for each model.

Consistent with prior research, gender differences in social cognition and overt aggression were found. Boys tended to be less skillful than girls in encoding (B= -.06, p<.01), goal clarification (B= -.20, p<.01), and response selection (B= -.12, p<.01). They also had significantly higher means on teacher-rated overt aggression (B=.18, p<.01), controlling for pretest and all other covariates. Gender did not significantly predict emotion regulation or hostile attribution, when controlling for pretest differences.

Ethnic differences were significant for two social-cognitive variables. African Americans, on average, made more hostile attributions of intent (B= .09, p<.05) and reported less benign – more aggressive – goals (B= -.08, p<.05) than European Americans. Latinos also tended to have higher levels of hostile attribution, compared to European Americans (B=.03, p<.05). Latinas were less likely to set friendly goals (B= -.06, p<.05) than non-Latinas.

Table 15: Fixed and Random Effects for Models Estimating Main Effects

	Emotion Regulation		Encoding		Hostile Attribution†		Response Selection†	
Fixed Effects	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept	1.12***	.20	.44***	.03	.21***	.21	.49***	.04
Pretest SIP skill	.62***	.04	.31***	.03	.27***	.03	.37***	.04
Male	-.16	.08	-.06**	.02	.00	.03	-.12**	.03
African American	-.14	.10	-.03	.02	.09*	.04	-.04	.04
Latino	.08	.06	-.00	.02	.07*	.03	-.04	.03
MC	.38***	.02	.03*	.01	-.04	.04	.07*	.03
MC+	.34***	.03	.01	.01	-.12**	.04	.14**	.03
Random Effects								
L1 Variance(e)	.344***		.025***		n/a		n/a	
L2 Variance(u11)	.179***		.003***		n/a		n/a	

	Overt Aggression				Goal Clarification			
	Males=1		Females=1		Males=1		Females=1	
Fixed Effects	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Intercept	.12***	.02	.30***	.06	.66***	.05	.45***	.05
Pretest	.69***	.08	.69***	.08	.30***	.05	.30***	.05
Gender	.18**	.05	-.18**	.05	-.20**	.04	.20**	.04
African American	.03	.04	.03	.04	-.08*	.04	-.08*	.04
Latino	-.03	.03	-.03	.03	-.06*	.03	-.06*	.03
MC	-.02	.03	-.24**	.05	-.01	.03	.13*	.05
MC+	-.05*	.02	-.23**	.05	.08**	.02	.18**	.04
Gender**MC*	-.22*	.07	-.22*	.07	-.14*	.06	-.14*	.06
Gender**MC*+	-.18*	.07	-.18*	.0	-.11*	.05	-.11*	.05
Random Effects								
L1 Variance(e)	.061***				.065***			
L2 Variance(u11)	.010*				.001			

Note. Dagger (†) indicates single-level model. Est = unstandardized estimate. SE=Standard Error. L1=Level 1 and L2=Level 2.
*p< .05. ** p<.01. *** p < .001

Children participating in the *MC* and *MC*+ interventions had better posttest ratings on social cognition and behavior, on average, than their same-sex peers in the comparison group. Both interventions affected emotion regulation, response selection, goal clarification, and overt aggression, however effects for encoding were found solely for *MC* and effects for hostile attribution were found solely for *MC*+.

Children in the *MC* cohort were better able to manage emotions (B=.38, p<.001), encode social cues (B=.03, p<.05), and select non-aggressive responses (B=.07, p<.05), than the comparison cohort. Boys and girls participating in *MC* were affected differently on two outcomes. *MC* boys had lower levels of overt aggression (-.24, p<.001) and reported more benign (e.g., friendly and neutral) social goals (.13, p<.05) at posttest, than comparison-group boys. However *MC* girls did not experience similar improvements, relative to comparison-group girls, on these outcomes.

Children in *MC*+ tended to experience somewhat stronger effects. Significant effects on emotion regulation (B=.34, p<.001), response selection (B=.14, p<.01), goal clarification, and overt aggression were also found for children in the *MC*+ cohort. In contrast to the pattern of effects found for *MC*, *MC*+ resulted in significant behavioral improvement for both genders, with boys experiencing greater effects (i.e., overt aggression: B_{girls}= -.05, p<.05; B_{boys} = -.23, p<.01; and goal clarification (B_{girls}= .08, p<.01; B_{boys} = .18, p<.01). Finally, children in the *MC*+ cohort had significantly lower levels of hostile attribution than children in the comparison group cohort (B= -.12, p<.001).

Effect sizes for program effects are presented in *Figure 7*. Effect size statistics are important because they provide a more meaningful interpretation of program effects and are comparable across studies. According to Coe (2000):

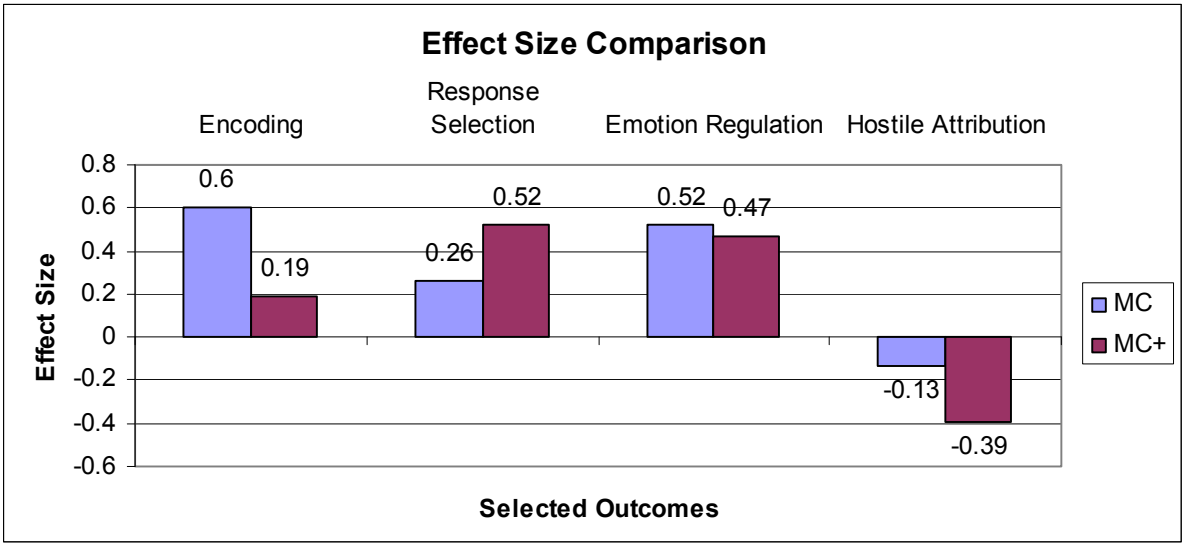
> An effect size is exactly equivalent to a 'Z-score' of a standard Normal distribution. For example, an effect size of .80 means that the score of the average person in the experimental group exceeds the scores of 79% of the control group.

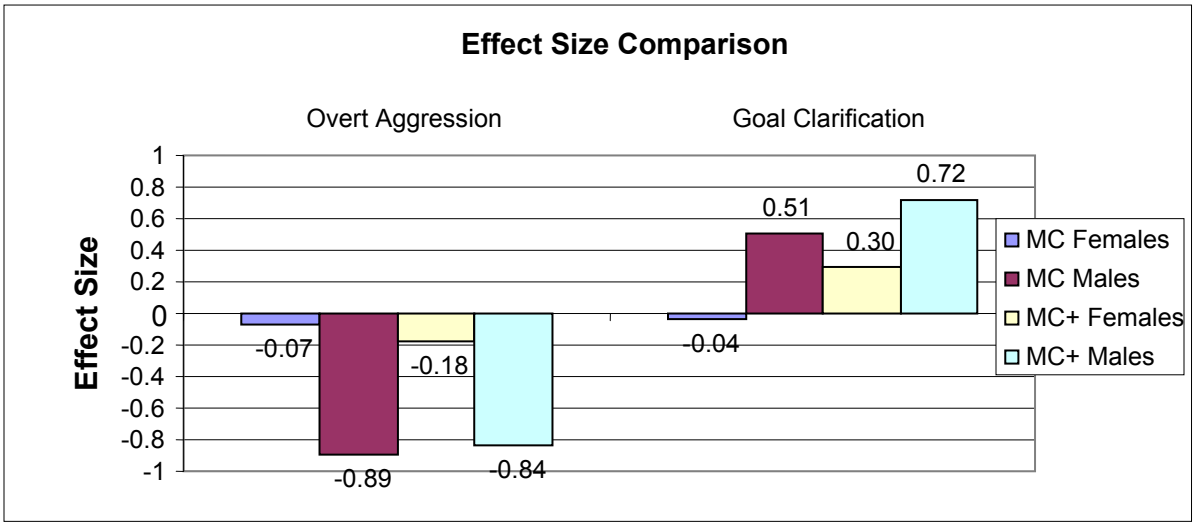
Effect sizes for *MC* and *MC*+ on theoretical mediators varied in magnitude from small to large (Cohen, 1988). According to Cohen (1988), effect sizes ranging from .20 to .49 are small, those from .50 to .79 are medium, and those from .80 to 1.0 are large in magnitude. *Making Choices* resulted in small to medium effects on emotion regulation, encoding, hostile attribution, and response selection, ranging in absolute value from .13 to .60, and *MC*+ had slightly larger effects on these mediators (i.e., $.30 \leq \delta \leq .52$).

Effect sizes for overt aggression and goal clarification varied by gender. For males, both interventions resulted in large effect sizes for overt aggression (i.e., overt aggression: δ_{MC}= -.89; δ_{MC+}= -.84). Effects on goal clarification were different in magnitude for *MC* and *MC*+, with a medium effect size for *MC* (δ_{MC}= .51) and a large effect size for *MC*+ (δ_{MC+}= .72). For females, small effect sizes were estimated for these two variables [i.e., δ_{MC}= -.07 (*ns*) and δ_{MC+}= -.18 (p<.05) for overt aggression, and δ_{MC}= .04 and δ_{MC+}= .30 for goal clarification].

On balance, effect sizes from this study were only slightly different from effect sizes obtained from Fraser et al. (2005). Comparing the previous study to the current study, the effects of *MC* were slightly weaker for hostile attribution (-.17 v. -.13), and slightly stronger for response selection (δ=.18 v. δ=.26). The effects of *MC*+ on response selection (δ=.54 v. δ=.52) and encoding (δ=.77 v. δ=.60) were slightly weaker. Larger differences in the effects of *MC*+ were found for hostile attribution (δ = -.55 v. δ= -.39) and for *MC* on encoding

Figure 7: Effect Sizes for MC and MC+ Programs





(δ=.82 v. δ=.60). Differential effects on SIP skill may have resulted from the use of a pretest-posttest (instead of a posttest-only) design, the use of modified measures, or from sample-based differences. Effects on overt aggression and goal clarification cannot be compared as these effects were not disaggregated by gender in the prior analysis.

Does Gender Moderate Program Effects?

Gender appears to moderate program effects on overt aggression and goal clarification. Gender differences in posttest means for these variables are displayed in figures

9 and 10. Program-by-gender interaction terms were either marginally-significant or non-significant for all other variables: emotion regulation ($B_{MC*male}$= .33, p<.10; $B_{MC+*male}$= .09, *ns*), and encoding ($B_{MC*male}$= .-.03, *ns;* $B_{MC+*male}$= -.01, *ns*), hostile attribution ($B_{MC*male}$= .-.05, *ns;* $B_{MC+*male}$= -.13, p<.10), and response selection ($B_{MC*male}$= .04, *ns*; $B_{MC+*male}$= .02, *ns*).

Consistent with study hypotheses, gender moderated the effects of *MC* and *MC*+ on overt aggression. Whereas boys in the *MC* and *MC*+ groups had significantly lower posttest scores on overt aggression than boys in the comparison group, girls in the *MC* cohort did not experience similar gains compared to their same-sex counterparts. *MC* boys had .24 lower posttest scores on overt aggression than comparison boys (p<.001), whereas *MC* girls had scores that were only .02 lower (*ns*) than comparison girls. Similarly, *MC*+ boys had .23 lower posttest scores on overt aggression than comparison boys and *MC*+ girls had scores that were .05 lower than comparison girls. Both of these effects were significant.

Although gender was not expected to moderate program effects on theoretical mediators, the main effects analysis found significant program-by-gender interaction effects for goal clarification. Males in the *MC* cohort experienced significant program gains in this outcome, while females in the *MC* cohort experienced no such effect. Gender-balanced effects were found for *MC*+, with higher levels of goal clarification skill obtained by both sexes; however, males experienced larger effects than females. On average, posttest means for goal clarification were .13 higher for boys in the *MC* cohort than comparison group boys (p<.001), whereas posttest means for girls in the *MC* cohort were .01 lower (*ns*) than comparison group girls. Similarly, posttest means for boys in *MC*+ were .18 higher than

posttest means for comparison group boys, whereas they were only .09 higher for girls in *MC+*, in relation to comparison group girls.

Figure 8: Program-by-Gender Interaction Effects for Overt Aggression

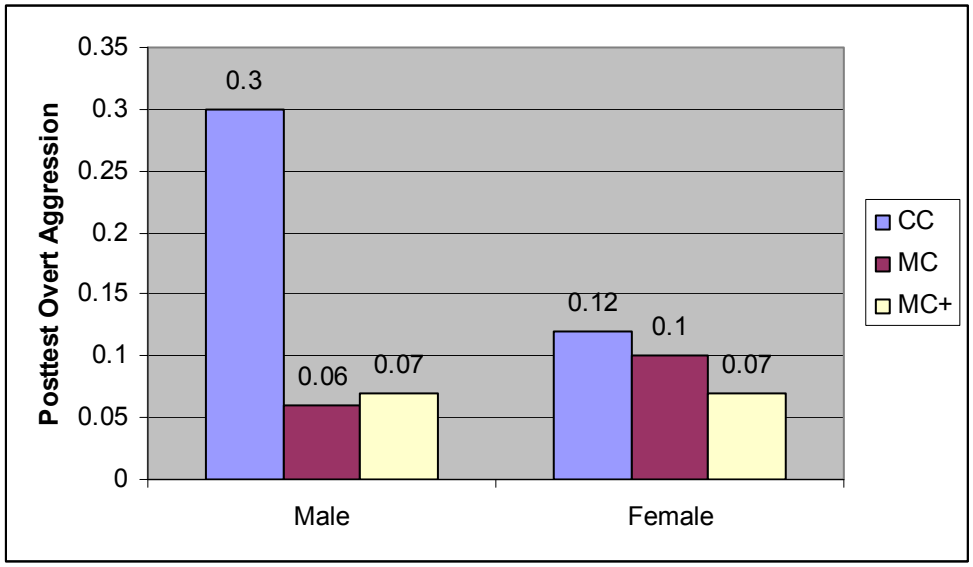
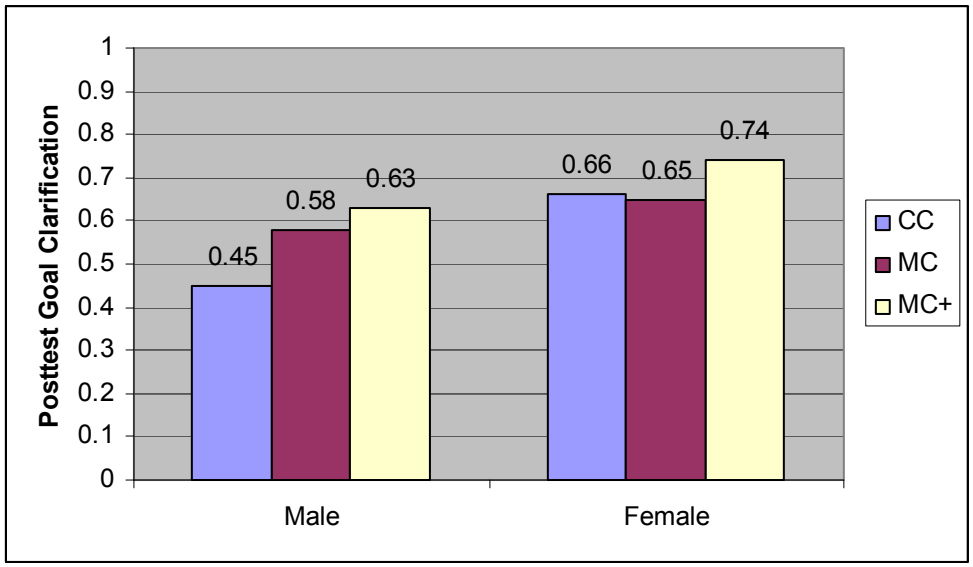


Figure 9: Program-by-Gender Interaction Effects for Goal Clarification



To test moderating effects of gender on the *b* path, an interaction term between gender and the theoretical mediator was constructed and added as a predictor to the full model

predicting overt aggression. However, the introduction of this term decreased the CFI to .70 and produced a negative TLI statistic. Thus, due to statistical issues created by the introduction of this interaction term, and due to the lack of a multiple group modeling approach, the hypothesis that gender would moderate the relationship between SIP and aggression could not be addressed in this analysis.

Do Theoretical Mediators Explain Program Effects on Overt Aggression?

Three out of five theoretical mediators appeared to explain some proportion of the effect of *MC* and *MC+* on overt aggression: goal clarification, response selection, and emotion regulation (see *Figure 10*). Although *MC* did not have significant effects on aggression in females, indirect effects of *MC* for both genders were estimated. This was carried out because indirect effects may be present, even in the absence of a significant main effect (Shrout & Bolger, 2002). As mentioned earlier, multilevel and single-level tests of mediation were performed, using multilevel structural equation modeling (MSEM) and general structural equation modeling (GSEM). Multilevel indirect effects were estimated for three out of five theoretical mediators (i.e., encoding, goal clarification, and emotion regulation). Single-level indirect effects were estimated for hostile attribution and response selection, because multilevel mediation models had a poor variance-covariance structure and would not converge.

In initial tests of mediation, posttest overt aggression was modeled as a latent variable with the mean as its indicator. The latent variable approach was used in order to remove measurement error in the dependent variable. Using the formula provided by McDonald (1999, p. 89), the factor loading was fixed to .91 (equal to $\sqrt{\omega}$) and the error term of mean was fixed to .17 (equal to 1- ω). However, this approach to estimating effects on overt

aggression was not used for the final mediation models, as it resulted in poor model fit. It is possible that measurement error in the dependent variable compromises model fit less than measurement error in the predictors, which is known to lead to inflated chi-square values and unreliable standard errors (Muthén, 2006).

To improve model fit and obtain more accurate parameter estimates, the mean of the overt aggression scale was used as the dependent variable. However, an issue with negative variance arose in the multilevel main effects and mediation models. Multilevel models revealed a negative group-level variance for overt aggression. An exploratory multilevel CFA revealed that Item 12 ('fights with others') had negative between-level variance that accounted for over 5% of the total variance. The item causing the negative variance was dropped, consistent with recommendations by Snijders and Boskers (1999). Negative variance of an indicator variable suggests that two indicators are highly correlated; the item with a negative variance is redundant with the other and can be dropped (C. Yang, personal communication, December 15, 2006).

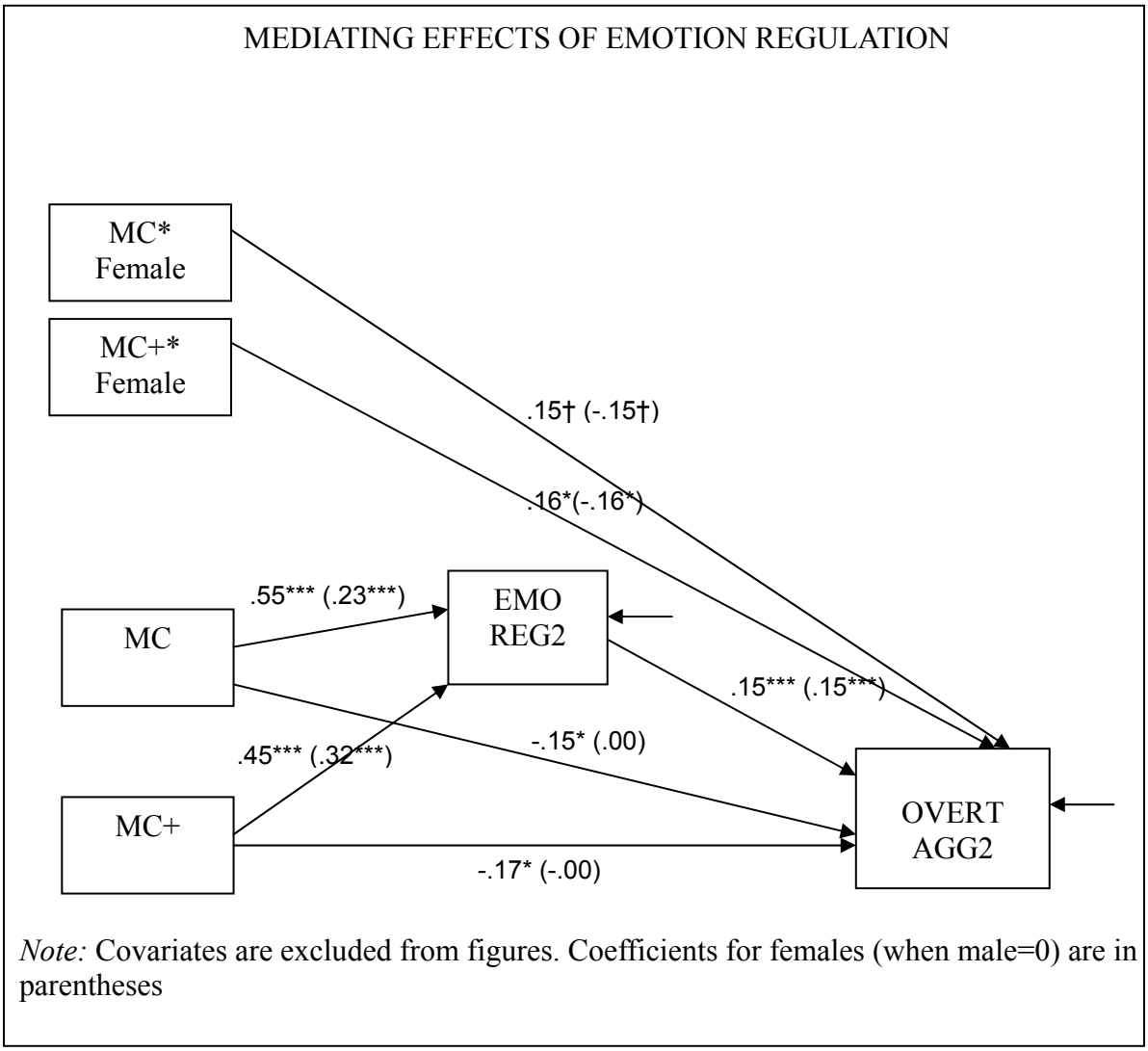
In general, indirect effects were estimated from mediation models that included program-by-gender interaction terms in the regression equations for the mediating and dependent variables. Moderated mediation could not be tested using this mediation model, as the addition of an additional interaction term (i.e., male*mediating variable) in the model predicting overt aggression led to poor model fit (negative TLI) and lack of model convergence. However, indirect effects were estimated separately for boys and girls, without testing the significance of gender differences (by using a multiple group approach) and without analyzing subsamples based on gender (see Methods chapter for details). The multilevel structural equation model used to estimate indirect effects is shown in *Figure 5*.

All indirect effects were tested using the Sobel Test (Sobel, 1982) and evaluated using the table of critical z values cited by MacKinnon et al. (2002, p. 90). The critical value listed for the smallest n (n=50) and a 2% cumulative frequency (p=.04), in the table listing cumulative frequency distributions for a dichotomous independent variable, was used (Z=.97). The results of these models are offered in *Figure 10*.

The strongest indirect effect was found for emotion regulation in boys (B_{MC}= -.09, p<.001 and B_{MC+} = - .07, p<.001) and in girls (B_{MC}= -.03, p<.05 and B_{MC+} = - .04, p<.01). In the female mediation model (CFI=.96, RMSEA=.13), the direct effects of *MC* and *MC*+ on overt aggression were totally mediated (i.e., they dropped to zero and became non-significant). In the male mediation model (CFI=.98, RMSEA=.09), direct effects dropped in magnitude, but remained significant, although significance levels decreased. Both the *a* paths (the program effects on emotion regulation) and the *b* path (the effect of emotion regulation on aggression) were significant at p<.001 (Females: B_{a1}=.23, B_{a2}=.32; Males: B_{a1}=.55, B_{a2}=.45; B_b=-.15), indicating that the program promoted this skill and that this skill was directly related to behavior improvement.

The model fit of the mediation model for goal clarification was also good (CFI=.99; RMSEA=.04). The results of this model suggest that program effects on goal clarification partially mediated program effects on overt aggression for boys and girls (see *Figure 10*). As in previous models, indirect effects appeared to vary by gender. Although goal clarification mediated program effects for both genders, effects of *MC* and *MC*+ were mediated for boys, but effects of *MC*+ only were mediated for girls (B_{ab}(*MC*)=.001, *ns*; B_{ab} B_{ab}(*MC*)=-.005, p<.05).

Figure 10: Significant Mediating Effects

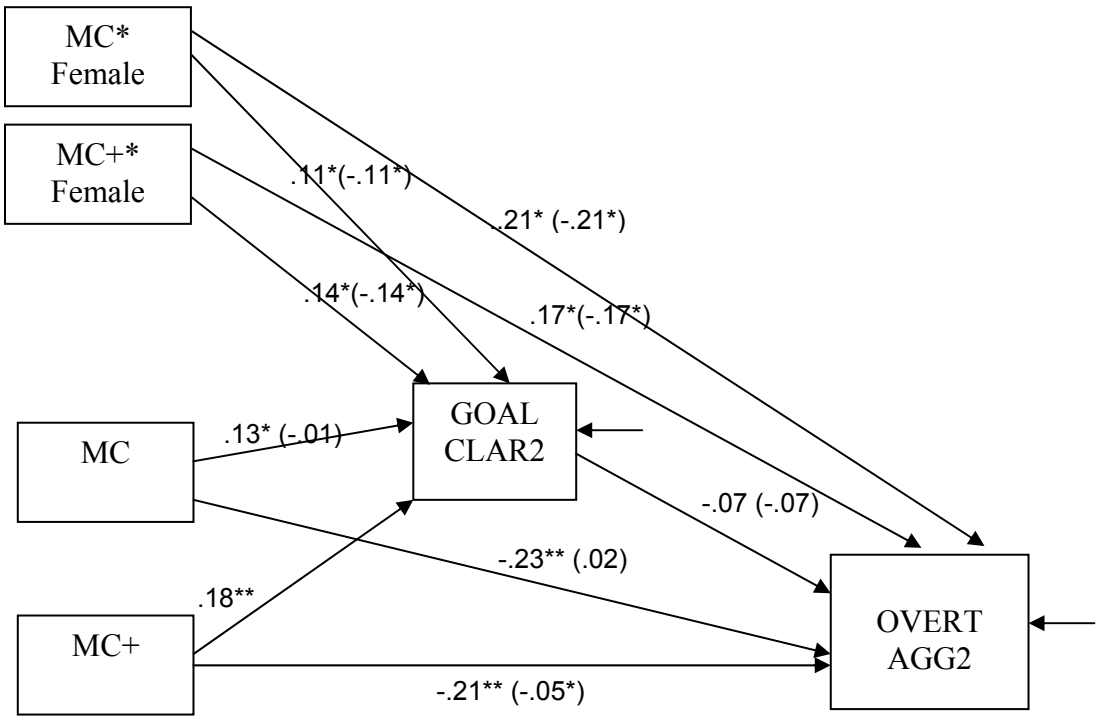


Note: Covariates are excluded from figures. Coefficients for females (when male=0) are in parentheses

Males	Females
Indirect Effect for *MC*= -.09, *p*<.001	Indirect Effect for *MC*= -.03, *p*<.001
Indirect Effect for *MC*+= -.07, *p*<.001	Indirect Effect for *MC*+= -.04, *p*<.01
CFI= .98; RMSEA= .09	CFI= .96; RMSEA= .13

Figure 10: Significant Mediating Effects (continued)

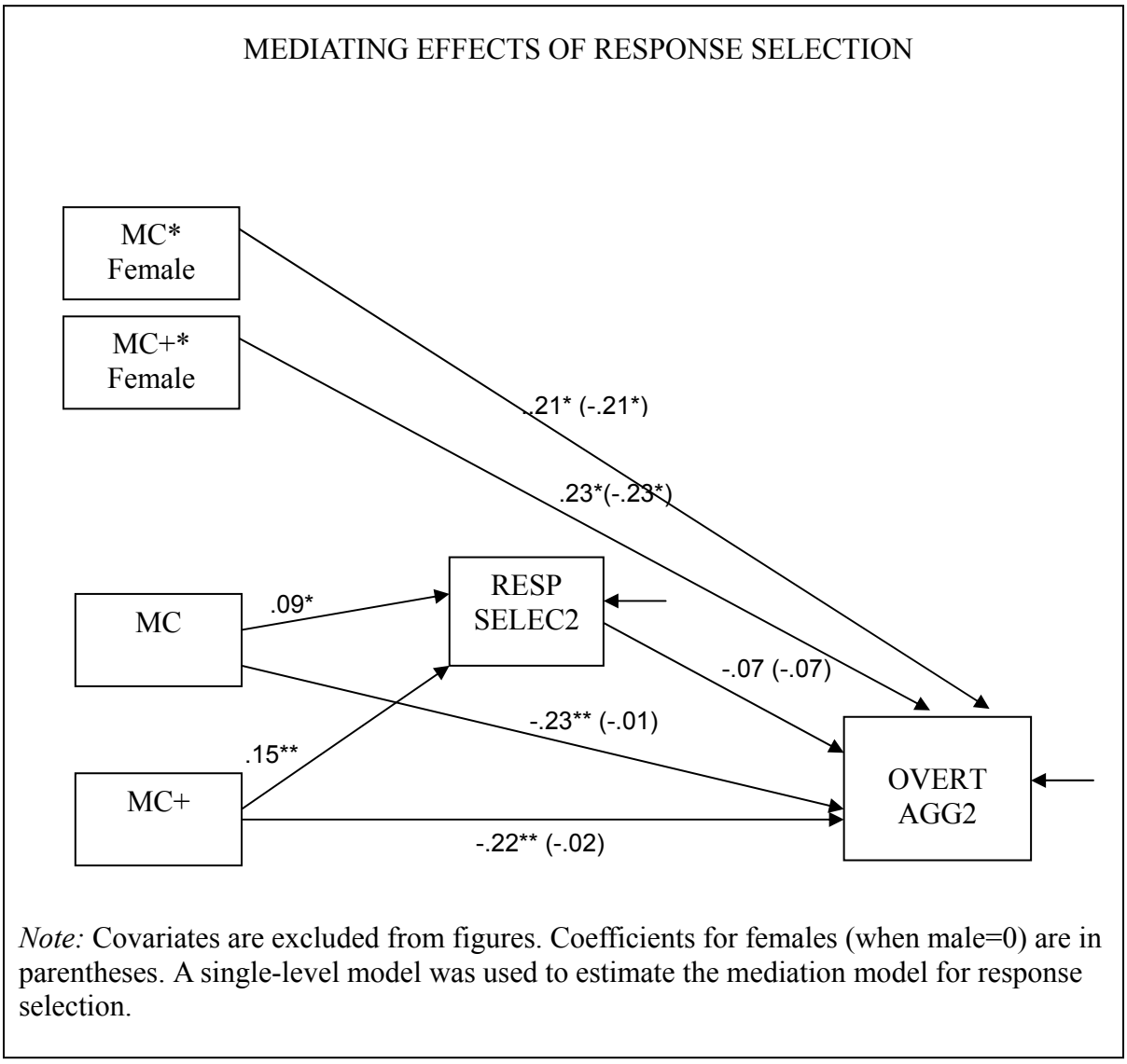
MEDIATING EFFECTS OF GOAL CLARIFICATION



Note: Covariates are excluded from figures. Coefficients for females (when male=0) are in parentheses

Males	Females
Indirect Effect for *MC*= -.004, $p<.05$	Indirect Effect for *MC*= .001, *ns*
Indirect Effect for *MC*+= -.008, $p<.05$	Indirect Effect for *MC*+= -.005, $p<.05$
CFI= .99; RMSEA= .04	CFI= 1.00; RMSEA= .00

Figure 10: Significant Mediating Effects (continued)



Note: Covariates are excluded from figures. Coefficients for females (when male=0) are in parentheses. A single-level model was used to estimate the mediation model for response selection.

Males	Females
Indirect Effect for *MC* = -.006, $p<.05$	Indirect Effect for *MC* = -.004, $p<.10$
Indirect Effect for *MC*+= -.011, $p<.05$	Indirect Effect for *MC*+= -.009, $p<.05$
CFI= 1.00; RMSEA= .00	CFI= 1.00; RMSEA= .00

Interestingly, *a* path coefficients were significant (see *Figure 10*) but the *b* path coefficient was not (B_b= -.07, *ns*). This may indicate that the significance of the indirect effect is driven by the effect on the theoretical mediator, rather than the effect of the mediator on the dependent variable. Although this is not consistent with the traditional Baron and Kenny (1986) criteria for mediation – e.g., significant effects for a) X→Y, b) X→M, and c) M→Y), it is consistent with current thinking about mediation and does not conflict with the finding of a significant mediating effect for this SIP skill.

Response selection also appeared to mediate program effects on overt aggression for males (B_{ab}(*MC*)= -.006, p<.05; B_{ab}(*MC*+)=-.011, p<.05) and females (B_{ab}(*MC*+)=-.009, p<.05), although the indirect effect of *MC* via response selection was only marginally significant for females (B_{ab}(*MC*)=-.004, p<.10). The fit statistics for both mediation models indicate perfect fit (CFI=1.0, RMSEA=.00). Component *a* path effects were significant for *MC* on males (Males: B_{a1}= .09, p<.05; Females: B_{a1}= .05, *ns*) and for *MC*+ on both genders (Males: B_{a2} = .15, p<.01; Females: B_{a2} = .13, p<.01). The effect represented by the *b* path was only marginally significant (B_b = -.07, p<.10). Given that the sample size for this test was larger than the sample size for emotion regulation and goal clarification, the non-significant *b* path is most likely not due to low power.

Encoding and hostile attribution did not mediate program effects, although significant program effects on these mediators were detected (significant *a* paths). An analysis of component effects found significant *a* paths (i.e., the effects of *MC* and *MC*+ on hostile attribution and an effect of *MC* on encoding) and found non-significant *b* paths (i.e., the effect of SIP variables on overt aggression).

Conclusion

The main effects of this study support the hypothesis that the *MC* and *MC*+ would result in social-cognitive and behavioral improvements. By and large, effect sizes were promising. As hypothesized, *MC*+ resulted in larger effect sizes than *MC.* The magnitude of program effects on boys' posttest overt aggression exceeded study expectations (i.e., δ_{MC}= -.82 and δ_{MC+}= -.89). Mixed support was found for study hypotheses relating to the moderating effect of gender, with gender moderating some effects and not others.

Finally, study hypotheses regarding mediation were generally supported by study findings. The majority of SIP skills (three out of five) partially mediated program effects on overt aggression. Response selection partially mediated the effect of both interventions for males, and it totally mediated the effect of *MC*+ for females. This may suggest that, for girls, response selection is more highly related to overt aggression. This hypothesis, however, could not be tested, due to the lack of a multiple group SEM approach. An alternative hypothesis is that total mediation (producing a non-significant direct effect, with the addition of a mediating variable) is easier to obtain for females, since the magnitude of the program effect on overt aggression is small to begin with.

Emotion regulation appeared to have the strongest mediating effect, by explaining the total effects of the program on overt aggression. Because emotion is most associated with reactive aggression, one might speculate that the program exerted impact on a reactive-overt form of aggression. However, this explanation is obfuscated by the fact that response selection and goal clarification – two constructs associated with proactive aggression (Crick & Dodge, 1996) – also mediated program effects. One possibility is that that these two variables are more not as highly related to proactive aggression during childhood as they are

during early childhood (Fontaine, Burks, & Dodge, 2002). Nonetheless, it appears that promoting emotion regulation is particularly important for controlling aggression in third grade.

Chapter VII

Discussion

Overview

Several findings of this dissertation research are worth highlighting. First, this study found that the *Making Choices* program promoted improvements in SIP skills. Although effects on these proximal outcomes varied, several medium and large effects were found. This finding is consistent with similar intervention research studies, suggesting that school-based prevention programs can successfully modify children's social-emotional skills and supporting the long-standing notion that social competence is not a trait-based characteristic, but a dynamic, environmentally-responsive set of social, emotional, and behavioral factors.

Second, the ability to regulate emotions, identify non-aggressive goals, and select benign responses to social problems appeared to explain behavioral improvement, validating the conceptual foundations of the program. Although the causality of this relationship cannot be established by the present analysis, the empirical literature appears to support an explanatory model, with SIP skill predicting the enactment of aggressive behavior and not the other way around.

Third, effect sizes for boy's overt aggression suggest that universal school-based interventions can indeed yield large effect sizes. This is a particularly important finding. It suggests that SIP skill deficits produce behavioral problems in high-risk and normative samples of children. More importantly, it contrasts results from Lipsey and Derzon's (1993)

meta-analysis of school-based violence prevention programs, which suggested that interventions conducted with children ages 6 to 10 result in small pretest-posttest effects on aggression (ES=.17; Lipsey & Derzon, 1993, p. 42).

Fourth, *MC* and *MC*+ had greater effects on overt aggression for boys than for girls. The moderating role of gender on these program effects can be explained in several ways. One explanation for differential effects on overt aggression relates to the measurement of aggression. It is possible that girls experienced similar decreases in aggression, but that these changes occurred for a more global form of aggression. In other words, although levels of physical aggression in elementary school girls are low, other forms of aggression (e.g., relational aggression, covert aggression, and oppositional-defiant behavior) may be more typical during this developmental period. Therefore, a broader construct of aggression may have captured more variation in girls' aggression. In the book entitled, *Aggression, Antisocial Behavior, and Violence among Girls*, Underwood & Coie (2004, p. 291) re-iterate the opinion expressed by Karen Bierman and her colleagues in an early chapter, which infers that aggression in females is a multifaceted construct:

> A broad screening strategy considering oppositional-defiant and attentional difficulties in addition to early aggression is more effective in identifying girls at risk for aggression and peer problems in the fourth grade and antisocial behavior in the seventh grade than is one that just targets physical aggression (Bierman et al., Chapter 7).

Common measures of aggression, such as the *CBCL* aggression subscale, include items relating to inattentive, disruptive, oppositional, overt, and covert aggression. However, because of the predictive utility of overt aggression, this study did not utilize the overall scale.

Another explanation for gender-driven moderating effects relates to the use of a universal sample. Normative samples of children commonly include a significant proportion of aggressive boys but include a much smaller proportion of aggressive girls. Thus, because girls, on average, exhibit low levels of overt aggression, non-significant program effects could be due to floor effects (i.e., girls' scores simply have less room to drop than boys' scores, because they have lower scores at baseline). Studies evaluating the effects of the *Fast Track* program with the high-risk sample appear to support this explanation. In these studies, high-risk girls and boys participating in a conduct problems prevention program (N=891; 31% female) experienced similar levels of behavioral improvement by the end of first and third grades (CPPRG, 1999, 2002). If additional studies were to show that high-risk boys experienced greater behavioral improvement from a gender-neutral prevention program than high-risk girls, then we could appropriately conclude that gender-specific interventions are needed to prevent conduct problems in girls. Unfortunately, high-risk, mixed-gender samples are in short supply in the field of prevention science. Lacking information about whether program-by-gender interaction effects occur in high-risk samples, researchers will continue to wonder whether gender-based variation in program effects is simply due to floor effects or due to fundamental differences in the etiology of aggression.

It is unclear why program effects on goal clarification were moderated by gender, while effects on other theoretical mediators were not moderated. Girls' self-reported social goals were more likely to remain stable from pretest to posttest, but boys' characterizations of their social goals were more likely to improve. One hypothesis is that girls characterize their social goals in less aggressive ways, due to socialization effects which inhibit the expression of aggressive intent. A second hypothesis relates to ceiling effects. Girls had high scores on

goal clarification at pretest and thus had little room to improve. Alternately, boys had lower goal clarification scores than girls, on average, so they experienced greater improvement.

Finally, indirect effects appeared to vary by gender. Emotion regulation and response selection totally mediated program effects on girls but only partially mediated program effects on boys. However, few studies have identified gender differences in the social-cognitive predictors of aggression and the SIP model. Lacking empirical evidence as to why this may be the case, we must turn to a statistical explanation. Program effects on overt aggression may have lost significance for girls and not for boys, because effects on girls were smaller in magnitude than effects on boys. Nonetheless, intervention research studies that examine whether certain social-cognitive skills are better than others at mediating program effects in boys versus girls are relevant and necessary.

This chapter will now turn to three main areas of discussion. First, strengths and limitations of this dissertation research will be reviewed, with particular attention to research design and statistical methods. Second, implications for practice are broadly discussed. The chapter concludes with recommendations for future research related to preventing aggressive behavior in children and adolescents.

Strengths

Several features of this dissertation research are noteworthy. First, this research involved a rigorous evaluation of program effects on five important proximal outcomes, four of which (i.e., encoding, hostile attribution, goal clarification, and response selection) had previously been evaluated using a posttest-only design in (Fraser et al., 2005) and one of which (i.e., emotion regulation) had not been previously evaluated. Evaluating proximal

outcomes contributes to our understanding of how this program affects social cognition and is consistent with a theory-based approach.

Second, the collection of data from a new cohort added rigor to the evaluation of main effects in two ways. First, the collection of pretest SIP data allowed the estimation of main effects on SIP skills, adjusting for levels of SIP skill at baseline. Controlling for pretest is critical when estimating effects of intervention, because children with lower levels of competence in certain areas tend to experience the greatest improvement. Second, all cohorts attended school while the *No Child Left Behind* policy was in effect. Thus, students in each condition experienced similar pressures by teachers, administrators, and parents to perform well on end-of-grade tests.

A third important characteristic of this research was its attempt to estimate mediating effects of targeted skills. Estimating the indirect effects of an intervention answers a critically important question: Did program effects on proximal outcomes lead to desired distal effects? Although many evidence-based programs exist, few evaluations of these programs have examined whether targeted skills account for program success (Gottfredson, 1998). Without this knowledge, one cannot justifiably conclude that a particular program produced effects due to the content that was delivered and not due to other factors related to program design, implementation, and school and community context.

Fourth, although multilevel structural equation modeling (MSEM) and hierarchical linear modeling (HLM) share much in common and, in some cases, yield identical estimates (Curran, 2003; Willett & Sayer 1994), a number of advantages are associated with MSEM, in performing tests of mediation. One advantage relates to the estimation of indirect effects. MSEM, by estimating a covariance matrix for the parameters in the indirect effect (*a* and *b*),

provides a more accurate test of indirect effects than HLM (S. Christ, personal communication, December, 15, 2006). HLM assumes this covariance to be zero, and, by doing so, introduces a potential source of bias in the estimation of the standard error for the indirect effect. Another advantage relates to the flexibility in model specification. In MSEM, the fixed and random effects do not have to be the same for each dependent or endogenous variable. Alternately, when testing mediation in HLM, fixed and random effects are uniform across every equation (K. J. Preacher, personal communication, April, 4, 2006).

A final advantage of MSEM, compared to HLM, relates to practical efficiency. MSEM allows the specification of "one model that describes all hypothesized relations between independent, intervening, and dependent variables" (Hox, 2002, p. 252). As a result, indirect effects and standard errors can be estimated easily by the software program, rather than being calculated by hand.

Fifth, although the use of multilevel modeling prevented the modeling of latent variables and the testing of moderated moderation by using a multiple group approach, there are advantages associated with the use of multilevel modeling. Multilevel models produce more accurate estimates and standard errors, when observations are nested within groups (or clustered). When data are nested, observations within each cluster tend to share common error variance (autocorrelation), violating statistical assumptions of uncorrelated error terms. Ignoring autocorrelation increases the likelihood of Type I error (rejecting the null hypotheses when, in truth, no true differences exist in the population; Krull & MacKinnon, 2001). Multilevel modeling addresses the problem of autocorrelation, by estimating random effects associated with a latent group-level intercept, in addition to fixed effects.

The primary limitation of using *Mplus* to estimate the multilevel path models – the inability to estimate random effects for more than two levels – did not affect the current study. The use of *SAS Proc Mixed* prior to the analysis verified that the data should be modeled using two-levels, without random slopes. Therefore, parameter estimates and standard errors should not be affected by an inadequate specification of random effects. Ultimately, a similar test of mediation could have been performed using a two-level, random intercept hierarchical linear model, but, after considering the advantages summarized above, MSEM was chosen as the preferred method.

Finally, there are several advantages of structural equation modeling (SEM) over multiple regression techniques that use Ordinary Least Squares (OLS) estimation. First, because SEM uses Maximum Likelihood (ML) estimation methods, it does not require normal variable distributions and balanced data. ML methods are more flexible in that they (a) attenuate bias associated with missing data; (b) produce standard errors that are robust to skewed variable distributions (Brown, 1984), and can be applied to unbalanced data (Curran, 2003). In addition, SEM is able to attenuate measurement error, by allowing the use of latent variables (though a latent variable approach was not used in the current study).

Limitations

This study is not without its limitations. For the purpose of simplification, these limitations have been organized into three categories: a) research design, b) measurement, c) statistical models, and d) model specification and significance testing.

Research Design

Several limitations of this study relate to research design issues. First, this study utilized a convenience sample; all third-grade classrooms in each school were invited to participate. The use of a convenience sample prohibits generalizability to the population that was sampled (i.e., children from rural, ethnically-diverse, low-to-middle-income backgrounds). Lack of external validity is a common problem in intervention research studies that employ non-random sampling strategies.

Second, participants were not randomly assigned to study conditions. Lack of randomization increases the potential for pre-existing differences between intervention and non-intervention participants. The presence of systematic differences between study conditions is a form of selection bias that can compromise the internal validity of a study. Selection bias can also result from sample and data attrition; this form of selection bias is called *sampling bias*. Sampling bias poses a threat to the internal validity of the study, when levels of attrition differ between treatment groups. Selection bias is a problem because it biases estimates of program effects (Rhodes, Pelissier, Gaes, Saylor, Camp, & Wallace, 2001).

While the use of a cohort design may have increased equivalence between study conditions, between-cohort comparisons on background measures revealed differences on several pretest measures. Using statistical controls mitigates selection effects but does not reduce endogeneity bias. More appropriate methods for addressing selection bias are: a) propensity score matching (Rosenbaum & Rubin, 1983); b); difference in differences (Heckman, Ichimura, & Todd, 1997); and c) matching with nonparametric regression (Fox, 2000). These methods control for selection on observed variables. Instrumental variable

methods may be used to control for selection on unobserved variables. However, these methods are most appropriate for quasi-experimental and non-experimental designs with large sample sizes. In the current study, the depletion of sample size resulting from the application of these methods would have compromised the estimation of indirect effects.

Third, although the collection of data from a new comparison group added rigor to the evaluation of main effects on SIP skills (by using a regressor-variable approach to control for pretest and by mitigating between-cohort differences due to history effects), it also introduced two complicating factors. One factor relates to the use of a treatment-withdrawal design. Because most teachers had participated in the intervention cohorts two and three years earlier, there may be unintended experimental contamination. However, spill-over effects are unlikely because, prior to data collection, teachers reported that they no longer utilized materials or activities from the *Making Choices* program. Another issue resulting from the lag in data collection relates to history effects. More specifically, during the two-year lag between the intervention and comparison cohorts, the ethnic composition of the school population changed, such that the percentage of Latino children significantly increased (e.g., 49% to 57% Latino over the course of 5 years). However, no differences between Latinos and non-Latinos on pretest aggression were found and there has been no precedence for differential effects based on race/ethnicity.

A fourth research design issue concerns potential rater effects. Two issues contribute to rater effects in this study: a) lack of triangulation between multiple informants and measures, and b) between-teacher differences in child ratings. Relating to triangulation issues, study variables were rated by only one source (either a teacher or a child); thus, inter-rater reliability could not be assessed. Furthermore, SIP skills were measured using only one

scale, so the convergent validity of the measure could not be checked. Between-teacher differences would not have been particularly problematic had all teachers rated all three cohorts, because rater effects would have influenced each cohort equally. However teacher turnover and teacher leave-of-absences increased the influence of raters on the measurement of program effects. For example, one teacher took a leave of absence and turned her classroom over to a new teacher, who taught during the second half of the year. In this case, the classroom was rated at pretest by one teacher and at posttest by another. Problems such as these are difficult to avoid when implementing a school- or community-based intervention, as the sociopolitical and economic context in which the intervention is delivered is subject to change at any moment.

Measurement

A number of measurement issues also affected the analysis. First, the low reliability of the hostile attribution measure may have resulted in attenuated effects for this variable. Low reliability also decreases the likelihood that findings will be replicated. Second, SIP scales tended to have low levels of concurrent validity. If the scales did not indeed accurately measure SIP skills, then the findings would need to be re-interpreted. We could not conclude that SIP skills were affected by the program and that they partially explained program effects on overt aggression. It may be that some other underlying construct is responsible for these effects. Further validation of this scale may be needed. Third, dropping items to increase the reliability and dimensionality of the scales is controversial. However, this strategy is seen as a viable way of reducing measurement error and increasing the integrity of the factor (Little, Lindenberger & Nesselroade, 1999; Bollen & Lennox, 1991). Finally, an item was dropped from the overt aggression scale due to negative variance. Dropping items is not common

practice, but is considered an appropriate strategy when an item or variable has a negative variance, as it is considered to be redundant in the scale (Bollen, 1989).

Statistical Models

The use of a multilevel approach to structural equation modeling led to three research issues: a) the inability to run multiple group structural equation models; b) lower power to estimate effects, due to lowered sample size; and c) the inability to utilize a latent variable approach, due to limited degrees of freedom.

Although multiple group SEM models are ideal for testing questions relating to moderation, the use of this method was prohibited by the use of a multilevel approach. In a multiple group analysis, the groups cannot be divided between different levels. Because the grouping variable was gender and the cluster was classroom (and classrooms included both males and females), it would be impossible to compare models for within- and between-level effects by gender.

The power analysis revealed the study had limited power to detect small effects. The use of multilevel models compromised the power of the study, by decreasing sampling units from the number of students (N=480) to the number of clusters (J=28). Smaller sample sizes also produced low power to test the significance of both direct and indirect effects. As mentioned in the methods chapter, the study had low power to detect small effects. Fortunately, effects were larger than expected and issues of power did not pose a major concern to significance testing.

Another issue related to the use of multilevel models relates to decreased sample size and degrees of freedom. When a latent-variable approach was attempted, the structural equation model had more parameters than degrees of freedom and the model could not be

identified. The latent variable approach would have maximized the benefit of using an SEM approach by differentially weighting items according to their factor loadings and removing measurement error from the variance of the factor.

Model Specification and Significance Testing Issues

The study was affected by issues related to testing mediation. The first issue relates to the lack of a longitudinal approach. In true mediation, there is temporal ordering between the independent variable, mediating variable, and dependent variable (MacKinnon et al., 2001). Although a significant mediating effect does not establish the presence of a causal mechanism (e.g., attributing change in the outcome variable to change in the mediating variable), it offers stronger evidence for drawing causal inferences (MacKinnon, Taborga, Morgan-Lopez, 2002). Therefore, the addition of an outcome variable measured at follow-up (i.e., Time 3 or later), would have improved the statistical test of mediation.

The second issue relates to the use of the Sobel Test in testing mediation (Sobel, 1982). Although the Sobel Test is one of the most popular ways of testing mediation, and is seen as a superior approach to testing mediation than the causal-steps approach (Baron & Kenny, 1986), it is generally more effective at testing indirect effects for studies with large samples (Mackinnon et al., 2002). In addition, the Sobel Test assumes that the distribution of the joint effect is normal; yet, in most cases, especially when sample size is small, the distribution of the joint effect is non-normal (Preacher & Hayes, in review). In such instances, bootstrapping methods are preferred (MacKinnon et al., 2002; Shrout & Bolger, 2002).

Finally, for the sake of comparison to the intent-to-treat study (Fraser et al., 2005), this study excluded covariates for dosage. In a recent unpublished analysis, child attendance

rates moderated program effects (Fraser, Rose, Terzian, & Guo, 2004). Therefore, heterogeneity in program effects caused by varying levels of exposure to the program, were not controlled. Future evaluations of the program may warrant the use of an efficacy subset approach.

Implications for Practice

Children with high levels of cumulative risk are likely to develop academic, behavioral, and social problems in adolescence and experience continued difficulties in adulthood. Prevention and intervention directed toward children at varied levels of risk can interrupt antisocial pathways and reduce the prevalence of conduct problems in childhood and adolescence. According to Walker et al. (1996), a universal primary prevention program is likely to prevent problem behaviors in approximately 75% to 85% of students participating in the program. This finding is promising and suggests that universal programs can help the majority of youths. This implies that secondary and tertiary programs are needed for 15% to 25% of children who are at greater risk or already exhibiting problem behaviors.

The current study suggests that helping children to identify and manage their emotions may be particularly important when delivering violence prevention programs to reduce aggression in middle childhood. Indeed, emotion-focused programs such as *PATHS* and *Second Step* have had much success. Other factors leading to the program's success may have been: a) the effective use of a manualized curriculum; b) the use of teacher and administrative feedback; c) adequate teacher training; and d) the provision of regular clinical supervision. All of these elements are considered critically important in the delivery of school-based interventions (Aber, Jones, Brown, Chaudry, & Samples, 1998; Galinsky & Terzian, 2006; Gottfredson & Gottfredson, 2001).

The stronger impact of *MC*+ relative to *MC* suggests that involving families in elementary school-based interventions may strengthen the efficacy of the intervention. From an ecological-developmental perspective, the involvement of families should increase the likelihood that program effects on behavior and social cognition will be sustained over time. During childhood, children are influenced by caregivers more so than during any other developmental period. Thus, it makes sense that involving caregivers, even if only minimally, can increase the capacity of school-based prevention programs to produce desired improvements in elementary-school youths.

Evidence-based preventive interventions hold the potential to disrupt negative developmental trajectories associated with childhood aggression and relieve some of the burden on systems involved in the education, care, and rehabilitation of youth. A number of 'promising' and 'model' violence prevention programs have been identified by national agencies and clearinghouses such as the Center for the Study and Prevention of Violence (CSPV; http://www.colorado.edu/cspv/blueprints/) and the Campbell Collaboration (http://www.campbellcollaboration.org/index.asp), each of which employ rigorous standards for evaluating intervention research studies.

Although the accumulation of this knowledge has contributed significantly to violence prevention efforts, even the most established evidence-based programs become less effective when implemented by school-based practitioners facing "real-world" pressures and constraints (Ozer, 2006). To facilitate the effective implementation of evidence-based interventions, it is important to identify critical program components and/or processes. Tests of mediation can inform the implementation and delivery of school-based programs, by helping to identify core ingredients of the intervention. Just as medications fail to work when

certain active ingredients are missing, interventions may be more likely to fail when particular areas of program content are not sufficiently covered. Four different types of intervening variables may be examined: a) process characteristics (e.g., participation of teachers and administrators in the development of the intervention); b) implementation characteristics (e.g., dosage, duration, fidelity); c) design characteristics (e.g., multicomponent, multielement, manualized), and d) program content (e.g., social-cognitive skills, life skills, character-development).

The current study explored whether program effects on overt aggression could be attributed to program content, however understanding process and implementation characteristics is just as important. Classroom, school, community, and political contexts exert powerful effects on program success. Factors such as effective disciplinary practices, classroom norms favoring aggression, principal support, and school resources have been tied to greater program success (Ozer, 2006). Implementation quality, e.g., factors related to dosage, duration of intervention, and treatment fidelity, also relate to program effectiveness (Kam et al., 2003; Wilson, Lipsey, & Derzon, 2003). More research exploring these effects is needed.

Finally, practitioners must know which interventions work for whom. Testing how individual characteristics, such as race, gender, and risk status influence program effects can inform practitioners in the process of program implementation and selection. Tests of moderation offer information about which programs elements may be more relevant for particular populations or subgroups of children. For example, interventions that promote self-esteem and the development of friendships may be more effective at reducing risk for girls, whereas interventions that teach alternate thinking strategies and promote empathy may be

more effective at reducing risk for boys. Equipped with knowledge about "what works" and "for whom," intervention researchers may be able to have a larger impact on a greater number of youths.

Recommendations for Future Research

Based on knowledge culled from the literature and from dissertation findings, recommendations for two areas of research – developmental psychopathological and intervention research – are provided. Developmental psychopathological research seeks to understand the epidemiology and etiology of problem behavior. Intervention research seeks to interrupt maladaptive developmental trajectories, by promoting personal and/or environmental resources and/or minimizing risks associated with problem behavior. These kinds of research work in tandem and provide important information for federal, state, and local crime prevention policy.

Recommendations for Developmental Psychopathological Research

Basic research in epidemiology and developmental science is needed with regard to several three different issues:

- articulate SIP-related risk mechanisms to better inform interventions;
- identify gender sensitive and gender-specific factors implicated in the development of aggression;
- examine the nature, course, and consequences of female aggression.

Articulate SIP-based risk mechanisms. The SIP model has gathered much support in the developmental psychopathological research on aggression. SIP skills are seen to play a key role in the likelihood that aggressive behavior will or will not be expressed. Although

there is strong evidence for bivariate relationships between SIP and aggression, we do not yet fully understand the interrelationships between components of SIP model (e.g., latent mental structures, physiological arousal, online SIP skills) and how these components assemble in meaningful ways to increase the likelihood that child with exhibit aggressive behavior. For example, the effect of emotion regulation and emotional understanding on patterns of social information processing has not been given adequate attention in empirical research. Articulating clear risk mechanisms and processes are a critical component of intervention research (Fraser, 2004). Tests of mediation using longitudinal data can help us to develop better conceptual models to inform preventive intervention.

Identify gender-sensitive and gender-specific factors. Thus far, research in developmental psychopathology has taken a gender-neutral approach in identifying risk factors for aggression. As a result, we lack knowledge about gender-specific and gender-sensitive factors for problem behavior in girls. Factors whose effects are moderated by gender may be referred to as gender-sensitive factors. In contrast, gender-specific factors – those risk and protective factors that uniquely affect the development of girls or boys – must also be identified. In a sense, these may be factors whose effects are mediated by gender. Variable-centered approaches with mixed-gender samples can help to identify whether gender moderates the impact of different risk and protective factors on aggressive, antisocial behavior. Growth mixture modeling with female samples can help us to understand what factors predict trajectories of antisocial behavior in girls. Finally, qualitative studies of violent juvenile and adult female offenders with retrospective data can provide a more complex understanding of the etiology and epidemiology of antisocial behavior in females.

Examine female aggression. In light of preliminary evidence supporting differences in the etiology of different subtypes of aggression (Kempes, Matthys, de Vries, & van Engelend, 2005; Vitaro, Brendgen, & Barker, 2006), more research examining the characteristics, pathways, and consequences of different forms of aggression (e.g., relational and physical; proactive and reactive) is needed. Literature on aggressive, antisocial in females is still in its nascent stages (Putallaz & Bierman, 2004). Exploratory person-centered analyses, such as latent class analysis and growth mixture modeling, can help to identify different types of aggression and describe trajectories of female aggression. Epidemiological research on female samples can facilitate the identification of different developmental patterns in the prevalence and nature of aggression over time.

Several studies have begun to identify trajectories of antisocial behavior (Silverthorn & Frick, 1999; Miller-Johnson, Malone, & the Conduct Problems Prevention Research Group, in preparation), however more are needed. Combined with person-centered analyses, longitudinal variable-centered analyses can help researchers develop gender-specific etiological models of aggression and examine within-gender heterogeneity (e.g., evaluating differences by SES, neighborhood context, and race/ethnicity). Basic research findings can be translated into applied research studies that test treatment protocols in a controlled setting, leading to more complex models of developmental psychopathology and the development of gender-specific and gender-sensitive interventions. All of these efforts would advance prevention science, by increasing the depth and breadth of intervention effects.

Recommendations for Intervention Research

To effectively prevent aggressive behavior, preventive interventions must target known developmental risks and promote protection (Coie et al., 1992; Fraser & Terzian,

2005). Promoting SIP skill and emotion regulation seems to be a worthwhile goal for elementary school-based prevention, in light of empirical evidence linking social-cognitive deficits to aggression in children. In addition to targeting known factors, preventive interventions seeking to reduce aggression must try to achieve the following objectives:

- estimate main effects on aggression at posttest and at one or more follow-up points (Durlak & Wells, 1997; Gottfredson, 1998);
- estimate main effects on proximal, theory-based mediators (e.g., social-emotional skills; Birckmayer & Weiss, 2000);
- calculate the indirect effect(s) of the program via proximal, theory-based mediator(s) and conduct more accurate tests of mediation (CPSV, 2006; Lipsey, 1988; Petrosino, 2000);
- explore how program effects vary by subgroup (e.g., moderating effects; Tein et al., 2004);
- avoid selection bias and low statistical power (CSPV, 2006);
- utilize analytic methods that account for clustering (Bloom ,2005; Gottfredson, 1998); and
- address dosage effects by controlling for dosage or utilizing an effect of treatment on the treated (ETT) approach (Heckman & Robb, 1985).

Preventive interventions must do a better job of modifying key behavioral risk factors, in order to improve children's academic functioning and social relations and prevent later drug use and delinquency. Modifying correlates aggression may yield some benefits to intervention participants, but this strategy alone is not sufficient. Explanatory risk factors must be mitigated, in order to disrupt risk mechanisms.

Testing whether proximal outcomes explain behavioral change is one way to identify key risk factors. The analysis of mediating factors is an important selection criterion for being identified as a Blueprints Model Program (CSPV, 2006):

> The Blueprints Advisory Board looks for evidence that change in the targeted risk or protective factor(s) mediates the change in violent behavior… In its reviews of different programs, the Advisory Board has discovered that many programs reporting significant deterrent "main effects" have not collected the data necessary to complete an analysis of mediating factors.

The handful of intervention studies (Bierman et al., 2002; Lochman & Wells, 2002; Spoth et al., 1998) that test mechanisms of program-induced change, tend to ignore the question of whether various subgroups (e.g., girls versus boys; low-risk versus high-risk) are affected via similar mechanisms. For instance, although a study may conclude that the intervention reduced aggression in boys and girls, and that changes were explained by proximal intervention targets, this information alone is insufficient. It is necessary to discern whether mediating factors and/or mechanisms vary by gender, given evidence for gender-specific etiologies for psychiatric disorder and problem behavior (Putallaz & Bierman, 2004) and given the call for gender-specific policies and programs (Bloom, Owen, Deschenes, & Rosenbaum, 2002). One way to test whether program effects vary according to the population involved is to conduct multiple-site replication (CSPV, 2006). According to CSPV (2006):

> Replication is an important element in establishing program effectiveness and understanding what works best, in what situations, and with whom. Some programs are successful because of unique characteristics in the original site that may be difficult to duplicate in another site (e.g., having a charismatic leader or extensive community support and involvement).

However, additional means of testing moderated mediation are needed.

Findings from this dissertation study suggest that males had greater decreases in aggression at posttest than females, a pattern of findings that has been found in similar intervention research studies. Considering recent increases in the rate and severity of female delinquency (Office of Justice Programs, 1998), it is imperative that we develop a better understanding of how best to reduce aggression in girls. To do this, we must discern whether risk factors and mechanisms involved the development of aggressive behavior are moderated by gender. We must also validate measures of aggression with female samples. Different forms of aggression, each having distinct etiologies (Vitaro, Brendgen, and Barker, 2006), may warrant different intervention strategies. Conversely, the same form of aggression may be associated with distinct etiologies for different subgroups.

Research examining the moderating effect of cumulative risk is also needed. Many programs have noted that intervention participants with higher baseline risk levels fare better than those with lower baseline risk levels. Yet, in many intervention research studies, the construct of risk is narrowly defined by teachers' ratings on a specific measure (such as aggression or peer rejection). Developing a more ecologically-valid measure of cumulative risk would provide a more solid determination of whether children at high-risk for poor developmental outcomes are truly able to benefit from universal intervention, or whether they are more likely to benefit from a two-pronged, multi-element approach to intervention, which includes indicated and universal components and targets multiple social domains (CPPRG, 1999a). For example, risk measures that account for factors related to social inequality and injustice, family and/or neighborhood violence, and socioeconomic status may provide a more valid measure of risk exposure.

Without question, intervention research that attempts to answer more complex questions carry the potential to advance our understanding of how increasing protection can reduce risk in the context of real-world challenges and situations. Intervention researchers can contribute to the field of prevention research, by estimating main effects on proximal and distal outcomes, testing explanatory mechanisms, and exploring whether different subgroups follow similar processes. With this knowledge in hand, interventionists can develop more effective intervention strategies and practitioners and policymakers can make more well-informed decisions about which interventions work best for the youth they serve.

Appendix A: Empirical Studies Reviewed (N=16)

Program	Citation	Study Design	Sample	Gender moderating effects	Social-cognitive effects	Emotional effects	Mediating effects
Promoting Alternative Thinking Strategies (PATHS) (Universal; 4 studies)	Greenberg, Kusché, Cook, & Quamma (1995)*	Randomized, controlled; pre/post only	N=286; Ages 7-9; 58% Male; 58% White; 32% Black; 10% Other	Tested moderating effects of gender but none found.	---	Increased feelings vocabulary & emotional understanding	---
	Greenberg & Kusché (1998)*	Randomized, controlled; pre/post only (The 1- & 2-year f-up analyses did not qualify for further review, due to design issues)	N=57 (Deaf) Ages 5-12 47% Male 83% White; 17% Other	---	Improved perspective-taking, outcome expectancies, means-end thinking, reading emotional cues, & response selection	Improved emotional recognition skills	---
	Kam, Greenberg, & Walls (2003)	Not randomized; pre/post only	N=350 Ages 5-6 47% Male 79% Black; 21% Other 85% F-R lunch eligible	---	---	Improved emotional competence	---
	Kam, Greenberg, & Kusché (2004)*	Randomized, controlled; pre-, post-, 1-, & 2-year follow ups.	N=133 Ages 6-9 73% Male 66% White; 34% Other	---	Decreased generation of aggressive solutions	Decreased self-reported depression; Increased negative feelings vocabulary	---

Appendix A: Empirical Studies Reviewed (N=16) – continued

Program	Citation	Study Design	Sample	Gender Moderating effects	Social-cognitive effects	Emotional effects	Mediating effects
Second Step (Universal; 6 studies)	Orpinas, Parcel, McAlister, & Frankowski (1995)	Not randomized; pre-, post-, & 3-month follow up	N=223 Ages 11-12 64% Latino 36% Other 50% F-R lunch eligible	Reduced boys' self-reported aggression, but not girls'	Improved attitudes toward hostile behavior & knowledge about violence	---	---
	Grossman et al. (1997)*	Randomized, controlled; pre-, post-, & 6-month follow up	N=790 Ages 7-9 54% Male 79% White 21% Other	---	---	---	---
	McMahon, Washburn, Felix, Yaking, & Childrey (2000)	No untreated control group; pre-, post-, & 1-year follow up	N=109 Age 3-5 100% F-R lunch eligible	None found	---	---	---
	Taub (2002)	Not randomized; pre-, post-, & 1-year follow up	N=70 Ages 9-11 Schools were 37 and 40% F-R lunch eligible	---	---	---	---
	Van Schoiack-Edstrom et al. (2002)	Not randomized; pre/post only	N=714 Ages 12-14 49% Male	Improved girls' attitudes towards aggression, but not boys'	Improved social self-efficacy & attitudes toward aggression.	---	---
	McMahon & Washburn (2003)	No control group; pre/post only	N=156 Ages 11-14 36% Male 100% Black	---	Improved self-reported knowledge & skills	Improved self-reported empathy	Increased empathy linked to decreased aggression

Appendix A: Empirical Studies Reviewed (N=16) – continued

Program	Citation	Study Design	Sample	Gender moderating effects	Social-Cognitive effects	Emotional effects	Mediating effects
Attributional Program (Indicated; 1 study)	Hudley & Graham (1993)*	Randomized, controlled; pre/post only	N=101 Ages 9-11 100% Male 100% Black	Not applicable	Less hostile attribution & decreased endorsement of aggression	Less anger	Anger mediated by hostile intent attribution & anger & intent attribution together mediated program effects on aggression
Anger Coping Program (Indicated; 5 studies)	Lochman, Burch, Curry, & Lampron (1984)*	Randomized, controlled; pre-, post-, & 1-month follow up (4 treatment conditions)	N=76 Ages 9-12 100% Male 53% Black 47% White	Not applicable	Increased self-esteem – no effect on generating alternative solutions	---	---
	Lochman & Curry (1986)	No control group; pre/post only	N=20 Ages 9-12 100% Male	Not applicable	Increased self-esteem	---	---
	Lochman & Lampron (1988)	Not randomized; pre-, post-, & 7-month follow up	N=31 Mean age 11.7 years 100% Male	Not applicable	---	---	---
	Lochman, Lampron, Gemmer, Harris, & Wyckoff (1989)	Randomized controlled; pre/post only	N=32 Ages 9-13 100% Male	Not applicable	Improved perceived social competence	---	---
	Lochman (1992)	Not randomized; 3-year f-up study	N=145 Ages 12-15 100% Male	Not applicable	Increased self-esteem & social prob-solving skills	---	---

Appendix B: Evaluation Design and Study Outcomes for Selected Studies

Program	Citation/ Study Design	Sample Characteristics	Instruments *Request references from author	Study Findings		
				Soc-Cog	Emotion	Behavior
PATHS (Promoting Alternate Thinking Strategies) *Universal	Greenberg et al. (1995): Pre/post only	286 students; Grades 2-3; 4 schools; 33% spec ed; 58 % male; 58% White; 32% Black; 9% Other	Emotional measures: Kusché Affective Interview; questions about feelings (13 questions).	- - -	Increased positive [$F(1,282) = 21.5$, $p<.001$], & negative emotion words [$F(1,282)=49.9$, $p<.001$] & emot.unders $F(1,268)=7.5$, $p<.01$].	---
	Greenberg & Kusché (1998): Pre/post only; 1- and 2-yr f-up analyses do not qualify	57 hearing-impaired students; Grades 1-6; 11 classrooms; 83% White; 17% Other	Social-cognitive measures: Social Problem Solving Assessment Measure-Rev.; Matching Familiar Figures Test Emotional measures: Kusché Emotional Inventory (Inv.) Behavioral measures: *Teacher-rated*: *TRF*; Walker Behavior Problem Identific. Checklist; *parent-rated*: *CBCL*; Eyberg Child Behav. Inv.	A trend found for fewer errors on a task measuring impulsivity; $F(1,51)=3.4$, $p=.07$. Improved SIP-related skills.	Improved emotional recognition [$F(1,51) = 44.0$, $p<.001$] and reading of emotional labels [$F(1,51) = 75.6$, $p<.001$].	Increased parent-rated social competence [$F(1,41)=4.5$, $p<.05$], but did not decrease parent-rated externalizing symptoms.
	Kam et al. (2004): Pre-, post-, and 1- and 2-year f-ups	133 students; Grades 1-3; 7 schools; 73% male; 66% White; 20% Black; 14% Other	Social-cognitive measures: Social Prob-Solving Interview Emotional measures: Kusché Affective Interview; Children's Depression Inventory Behavioral measures: *TRF*; Teacher–Child Rating Scale	Marginally significant reduction in aggressive solutions; $F=2.8$, $p=.07$	Increased negative feelings vocabulary two-years later (*ES*=.54). Decreased depression (*ES*=.49).	Small effect for reducing teacher-rated aggression (*ES*=.18).
Second Step Program *Universal	Grossman et al. (1997): Pre-, post-, and 6-mo f-up	N=790 students; Grades 2-3; 12 schools; 49 class-rooms; 54% male; 79% White; 7% Black; 4% Latino; 10% Other	Behavioral measures: *Teacher-rated:* School Social Behavior Scales; *TRF*; *Parent-rated: CBCL*; Parent-Child Rating Scale; *Observer-rated*: Behavioral observations of 12 students per classroom (n=588).	---	---	Decreased phys aggression ($p=.03$) & increased non-aggressive behav ($p=.04$). Less classrm aggressn ($p=.03$) at f-up.

Appendix B: Evaluation Design and Study Outcomes for Selected Studies (continued)

Program	Citation/ Study Design	Sample Characteristics	Instruments *Request references from author	Study Findings		
				Soc-Cog	Emotion	Behavior
Attributional Program (AP) *Indicated	Hudley & Graham (1993) Random assignment to two attributional programs and a control group.	101 students 100% male; ages 9-11; 100% Black. Two schools in an urban low-SES area. School population is predominantly African American (80%-92%) and 30% qualified for free and reduced lunch)	Behavioral ratings: Teacher Checklist: Aggression (8 items) – reactive and proactive; Prosocial behavior (5 items); School performance (4 items). Peer nominations to obtain social preference and aggression scores.	Compared to boys in the two other conditions, boys in AP had: (a) Fewer hostile perceptions of intentionality; $t(19)=8.08$, $p<.001$; (b) Less endorsement of hostile behavior; $t(19)=3.01$, $p<.05$.	Compared to boys in the two other conditions, boys in *AP* had significantly less anger; $t(19)=5.75$, $p<.001$.	Compared to boys in the two other conditions, boys in *AP* had: (a) less observed negative verbal behavior in the peer provocation task $F(2,64)=5.01$, $p<.01$; and (b) less teacher-rated reactive aggression; $F(2,126)=3.76$, $p<.05$.
Anger Coping Program (ACP) *Indicated	Lochman et al. (1984)	76 students 100% male; ages 9-12; 53% Black; 47% White.	Social-cognitive measures: Perceived Competence Scale for Children (PCSC) Behavioral measures: Breyer's Behavior Observation Schedule for Pupils and Teachers (BOSPT; Breyer & Calchera, 1971); Missouri Children's Behavior Checklist (MCBC)	Boys in the anger-coping intervention groups had marginally significant gains in self-esteem– $F(1,72)=3.77$, $p<.10$ – than boys in the comparison group. No effects on total problem solving score, but proportion of alternatives involving inhibited aggression improved – χ^2 (1, N=76)=4.28, $p<.05$.	---	Boys in the *ACP* had less teacher-rated off-task disruptive behavior – $F(1,72)=5.37$, $p<.05$ – and parent-rated aggression – $F(1,68)=6.42$, $p<.05$ – than boys in a comparison group.

Appendix C: Carolina Child Checklist-Teacher Form

Student name _____Completed by: _____Date _____Circle one: Pretest Posttest

Part One: Below is a list of items that describe students. Please circle the number that best describes the student within the last month. Please answer all items as well as you can, even if some do not seem to apply to this student.

	never	rarely	sometimes	often	very often	always
Works well alone	0	1	2	3	4	5
Lies	0	1	2	3	4	5
Thinks before acting	0	1	2	3	4	5
Can give suggestions and opinions without being bossy	0	1	2	3	4	5
Mind wanders	0	1	2	3	4	5
Shows poor effort	0	1	2	3	4	5
Excludes other kids from peer group	0	1	2	3	4	5
Physically fights	0	1	2	3	4	5
Takes other's property	0	1	2	3	4	5
Uses physical intimidation with peers to get what he or she wants	0	1	2	3	4	5
Completes assignments	0	1	2	3	4	5
Pretends to harm or kill others in play	0	1	2	3	4	5
Teases classmates	0	1	2	3	4	5
Can calm down when excited or all wound up	0	1	2	3	4	5
Is helpful to others	0	1	2	3	4	5
Controls temper when there is a disagreement	0	1	2	3	4	5
Has trouble accepting authority	0	1	2	3	4	5
Has social contact with others	0	1	2	3	4	5
Stubborn	0	1	2	3	4	5
Harms others	0	1	2	3	4	5
Is liked by classmates	0	1	2	3	4	5
Easily distracted	0	1	2	3	4	5
Expresses needs and feelings appropriately	0	1	2	3	4	5
Friendly	0	1	2	3	4	5
Is disliked by classmates	0	1	2	3	4	5
Excludes other kids from games or activities	0	1	2	3	4	5
Learns up to ability	0	1	2	3	4	5
Initiates interactions with others	0	1	2	3	4	5
Plays with others	0	1	2	3	4	5
Eager to learn	0	1	2	3	4	5
Plays aggressively (rough)	0	1	2	3	4	5
Stays on task	0	1	2	3	4	5
Avoids social contact	0	1	2	3	4	5
Breaks rules	0	1	2	3	4	5
Very good at understanding other people's feelings	0	1	2	3	4	5
Breaks things on purpose	0	1	2	3	4	5
Hits others on purpose	0	1	2	3	4	5
Concentrates	0	1	2	3	4	5
Lies to make peers dislike a student	0	1	2	3	4	5

	never	rarely	sometimes	often	very often	always
Yells at others	0	1	2	3	4	5
Resolves peer problems on his/her own	0	1	2	3	4	5
Uses profanity	0	1	2	3	4	5
Tells peers he or she won't like them unless they do what he or she says	0	1	2	3	4	5
Works hard	0	1	2	3	4	5
Pays attention	0	1	2	3	4	5
Says mean things about others	0	1	2	3	4	5
Uses toy or imaginary weapons in play	0	1	2	3	4	5
Self reliant	0	1	2	3	4	5
Plays with prosocial peers	0	1	2	3	4	5
Accepted by prosocial peers	0	1	2	3	4	5

Part Two Below is a list of items that describe students. Please circle the number that best describes the student within the last month. Please answer all items as well as you can, even if some do not seem to apply to this student.

0 = Not True (as far as you know)				1 = Sometimes True			2 = Very Often True
0	1	2	Argues a lot	0	1	2	Disrupts class discipline
0	1	2	Defiant, talks back to staff	0	1	2	Screams a lot
0	1	2	Bragging, boasting	0	1	2	Showing off or clowning
0	1	2	Cruelty, bullying, or meanness to others	0	1	2	Demands must be met immediately, easily frustrated
0	1	2	Demands a lot of attention	0	1	2	Explosive and unpredictable behavior
0	1	2	Destroys his/her own things	0	1	2	Stubborn, sullen, or irritable
0	1	2	Difficulty following directions	0	1	2	Sudden changes in mood
0	1	2	Disobedient at school	0	1	2	Talks too much
0	1	2	Disturbs other pupils	0	1	2	Teases a lot
0	1	2	Easily jealous	0	1	2	Temper tantrums or hot temper
0	1	2	Gets in many fights	0	1	2	Threatens people
0	1	2	Physically attacks people	0	1	2	Unusually loud

Thank you for completing the Carolina Child Checklist.

Appendix D: Skill Level Activity Instrument

PRACTICE EXAMPLE

Example: On the playground.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN

CAN'T TELL

B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen? (Mark one answer)

- ☐ To have fun with your friend
- ☐ To get back at your friend for being mean

D. What would you do? (Mark one answer)

- ☐ Push your friend on the ground
- ☐ Ask, "What would you like to play?"
- ☐ Tell other people that Lou is mean

Story #1: Riding the bus.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN

CAN'T TELL

B. Look at the picture and circle all of the clues that tell you what is happening.

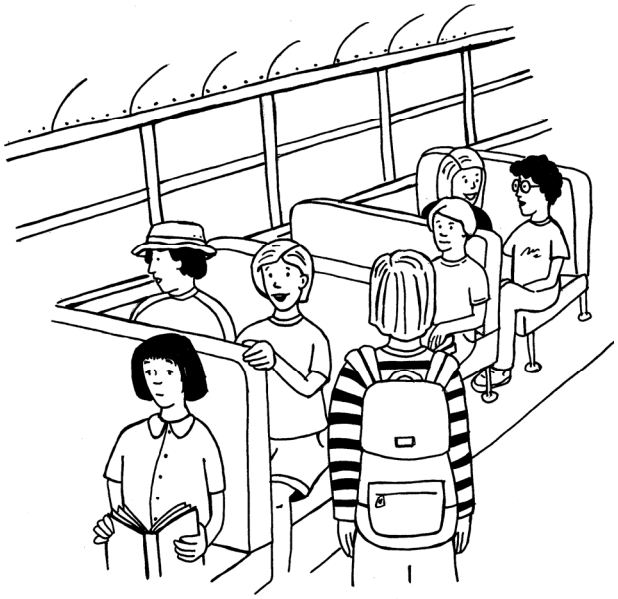
C. If you were the person in the story, what would you want to happen?

(Mark one answer)

- □ Just forget it and find an empty seat
- □ Make your friend sorry he didn't save a seat

D. What would you do? (Mark one answer)

- □ Hit Lee on the head
- □ Say, "I'll go look for a seat and talk to you later"
- □ Tell someone else, "Lee is a liar"



Story #2: In the cafeteria

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN

CAN'T TELL

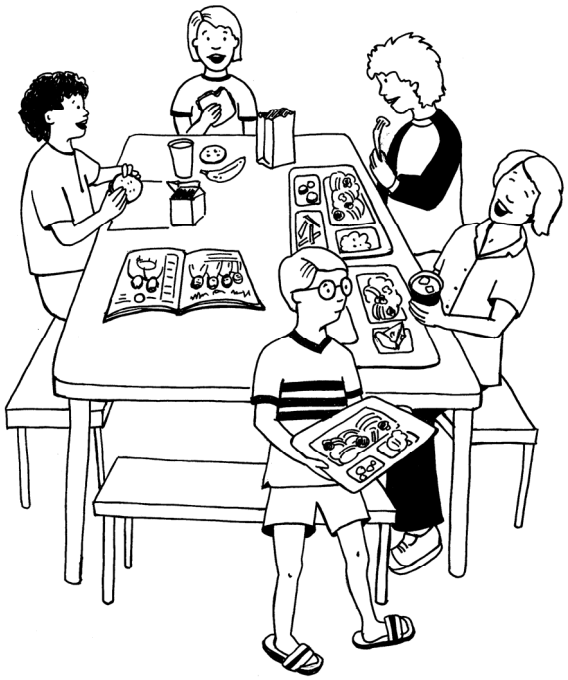
B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen first? (Mark one answer)

- ☐ To get back at those kids for laughing
- ☐ To eat lunch with your friends

D. What would you do next? (Mark one answer)

- ☐ Tell your friends not to play with those mean kids
- ☐ Pretend you didn't see them laughing
- ☐ Yell, "What are you laughing at!"



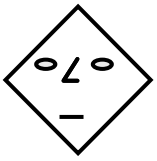
Story #3: At the pool.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN



CAN'T TELL

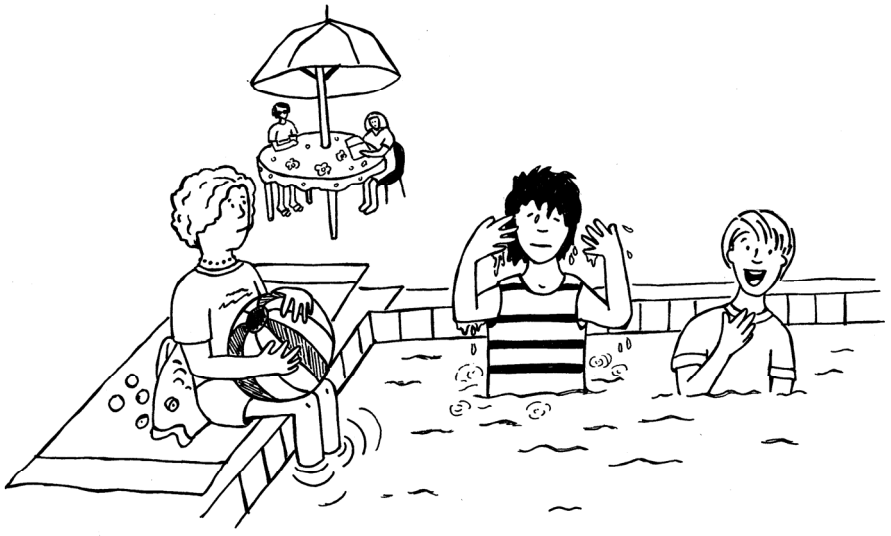
B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen? (Mark one answer)

- ☐ To hurt the kid
- ☐ To make sure the kid doesn't do it again

D. What would you do? (Mark one answer)

- ☐ Start a game without him in it
- ☐ Push him under the water
- ☐ Tell the kid to quit it



Story #4: Walking outside.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN

CAN'T TELL

B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen first? (Mark one answer)

- ☐ Go away and later ask Lou what was going on
- ☐ Get back at Lou

D. What would you do next? (Mark one answer)

- ☐ Plan how you would talk to Lou
- ☐ Plan your own party and tell your friends not to bring Lou
- ☐ Plan your own party and tell Lou he/she is not invited

Story #5: Your new magazine.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE

MEAN

CAN'T TELL

B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen first? (Mark one answer)

- ☐ Find someone else to show the magazine
- ☐ To make your friend sorry he/she did that

D. What would you do next? (Mark one answer)

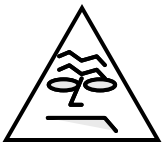
- ☐ Find your friend and ask what happened
- ☐ Find their backpack and throw it on the ground
- ☐ Tell your other friends that he/she can't be trusted

Story #6: New shoes.

A. Why did the person in the story do what she or he did? Draw an X on the face you choose.

FRIENDLY

MISTAKE



MEAN

CAN'T TELL

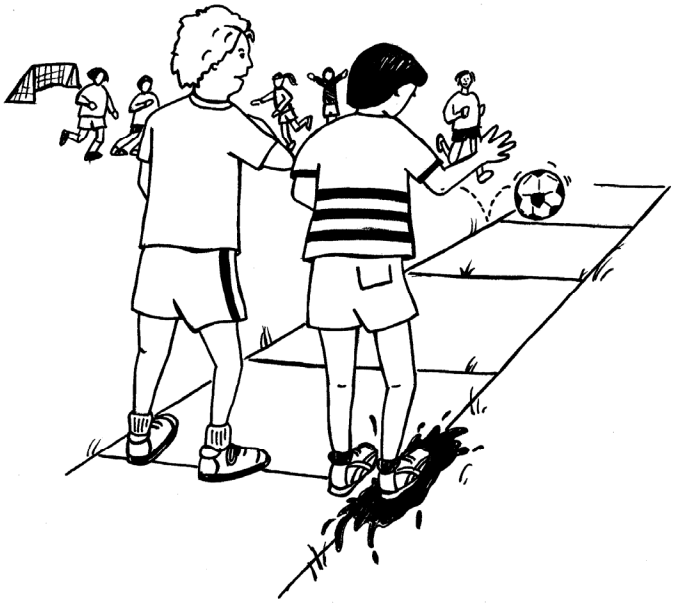
B. Look at the picture and circle all of the clues that tell you what is happening.

C. If you were the person in the story, what would you want to happen? (Mark one answer)

- ☐ Fight the kid
- ☐ Check your shoes

D. What would you do next? (Mark one answer)

- ☐ Push him/her back
- ☐ Tell everyone, "That kid is a bully"
- ☐ Clean your shoes



REFERENCES

Aber, J. L., Jones, S. M., Brown, J. L., Chaudry, N., & Samples, F. (1998). Resolving conflict creatively: Evaluating the developmental effects of a school-based violence prevention program in neighborhood and classroom context. *Development and Psychopathology*, 10, 187–213.

Abbot, R. D., O'Donnell, J., Hawkins, J. D., Hill, K. G., Kosterman, R., & Catalano, R. F. (1998). Changing teacher practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry, 68*, 542-552.

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32,* 328-34.

Achenbach, T. M., & Edelbrock, C. S. (1991). *Manual for the child behavior checklist and revised child behavior profile.* Burlington, VT: University Associates in Psychiatry.

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

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology, 20,* 93-114.

Arsenio, W, & Lemerise, E. (2001). Varieties of childhood bullying: Values, emotion processes, and social competence. *Social Development, 10,* 59-73.

Asher, S.R., & Hymel, S. (1981). Identifying children at risk in peer relations: A critique of the rate-of-interaction approach to assessment. *Child Development*, *52*, 1239-1245.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis, 2*, 119-124.

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., et al. (2002). Using the Fast Track randomized prevention trial to test the early-starter model of the development of serious conduct problems. *Development and Psychopathology, 14*, 925-943.

Bierman, K. L., Smoot, D. L., and Aumiller, K. (1993). Characteristics of aggressive-rejected, aggressive (non-rejected) and rejected (non-aggressive) boys. *Child Development, 64*, 139-151.

Bierman, K. L., & Wargo, J. B. (1995). Predicting the longitudinal course associated with aggressive-rejected, aggressive (nonrejected), and rejected (nonaggressive) status. *Development and Psychopathology, 7,* 669–682.

Birckmayer, J. D., & Weiss, C. H. (2000). Theory-based evaluation in practice: What do we learn. *Evaluation Review, 24*, 407-431.

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *British Medical Journal, 314,* 572.

Bloom , B., Owen, B., Deschenes, E. P., & Rosenbaum, J. (2002). Moving toward justice for female juvenile offenders in the new millennium: Modeling gender-specific policies and programs. *Journal of Contemporary Criminal Justice, 18,* 37-56.

Bloom, H. S. (2005). *Learning more from social experiments: evolving analytic approaches*. Newbury Park, CA, Sage.

Bohnert, A. M., Crnic, K. A., Lim, K. G. (2003). Emotional competence and aggressive behavior in school-age children. *Journal of Abnormal Child Psychology, 31,* 79-91.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. A. & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305-314.

Brown, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62-83.

Breyer, N. L., & Calchera, D. J. (1971). A behavioral observation schedule for pupils and teachers. *Psychology in the Schools, 8,* 330–337.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen and J.S. Long (Eds.), *Testing Structural Models* (pp. 136-162). Newbury Park, CA: Sage Publications, Inc.

Burks, V.S., Laird, R.D., Dodge, K.A., Pettit, G.S., & Bates, J.E. (1999) Knowledge structures, social information processing, and children's aggressive behavior. *Social Development*, 8, 220-236.

Burns, B. J., Costello, E. J., Angold, A., Tweed, D., Stangl, D., Farmer, E. M. Z., et al. (1995). Children's mental health service use across sectors. *Health Affairs, 14,* 147-159.

Burns, B. J., Phillips, S. D., Wagner, H. R., Barth, R. P., Kolko, D. J., Campbell, Y., & Landsverk, J. (2004). Mental health need and access to mental health services by youths involved with child welfare: A national survey. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 960-970.

Butler, L. & Meichenbaum, D. (1981). The assessment of interpersonal problem-solving skills. In P. Kendall & S. Hollon (Eds.), *Assessment strategies for cognitive-behavioral interventions*, (pp. 197-225). New York: Academic Press.

Cadesky, E. B., Mota, V. L., Schachar, R. J. (2000). Beyond words: How do children with ADHD and/or conduct problems process nonverbal information about affect? *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 1160-1167.

Campos, J. J., Mumme, D. L., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. In N. A. Fox (Ed.), The development of emotion regulation. Biological and behavioral considerations (pp. 284-303). *Monographs of the Society for Research on Child Development, 58* (2-3 Serial No. 240).

Camodeca, M., & Goosens, F. A. (2004). Aggression, social cognitions, anger, and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry, 46,* 186-197.

Camodeca, M., Goosens, F. A., Schuengel, C., & Terwogt, M. M. (2003). Links between social information processing in middle childhood and involvement in bullying. *Aggressive Behavior, 29,* 116-127.

Carvajal, S. C., Baumler, E., Harrist, R. B., & Parcel, G. S. (2001). Multilevel models and unbiased tests for group based interventions: Examples from the Safer Choices study. *Multivariate behavioral research, 36,* 185-205.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Coie, J. D., & Dodge, K. A. (1998). Aggression and antisocial behavior. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology (5th ed.): Vol. 3. Social, emotional, and personality development* (pp. 779-862). New York: Wiley.

Coie, J. D., Lochman, J. E., Terry, R., & Hyman, C. (1992). Predicting early adolescent disorder from childhood aggression and peer rejection. *Journal of Consulting and Clinical Psychology, 60*, 783-792.

Coie, J. D., Watt, N. F., West, S. G, Hawkins, J. D., Asarnow, J. R., Markman, H. J., Ramey, S. L., Shure, M. B., & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48,* 1013–1022.

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

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

Conduct Problems Prevention Research Group (2002). Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology, 30,* 19-35.

Connor, D. F., Steingard, R. J., Anderson, J. J., & Melloni, R. H. (2003). Gender differences in reactive and proactive aggression. *Child Psychiatry and Human Development, 33,* 279-294.

Côté, S., Zoccolillo, M., Tremblay, R.E., Nagin, D., & Vitaro, F. (2001). Predicting girls conduct disorder in adolescence from childhood trajectories of disruptive behaviors. *Journal of the American Child and Adolescent Psychiatry, 40*, 678-684.

Courtney, M. L., & Cohen, R. (1996). Behavior segmentation by boys as a function of aggressiveness and prior information. *Child Development, 67*, 1034-1047.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development, 67,* 993-1002.

Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development, 73,* 1134-1142.

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69*, 1630-1639.

Cuddy, M. E., & Frame, C. (1991). Comparison of aggressive and nonaggressive boys' self-efficacy and outcome expectancy beliefs. *Child Study Journal, 21,* 135–151.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529-569.

Dearing, K. F., Hubbard, J. A., Ramsden, S. R., Parker, E. H., Relyea, N, Smithmyer, C. M., et al. (2002). Children's self-reports about anger regulation: Direct and indirect links to social preference and aggression. *Merrill-Palmer Quarterly, 48,* 308-336.

Delveaux, K. D., & Daniels, T. (2000). Children's social cognitions: Physically and relationally aggressive strategies and children's goals in peer conflict situations. *Merrill-Palmer Quarterly, 46,* 672–692.

Denham, S. A., Caverly, S., Schmidt, M., Blair, K., DeMulder, E., Caal, S., et al. (2002). Preschool understanding of emotions: contributions to classroom anger and aggression. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43,* 901-916.

Department of Health and Human Services (2004, March 1). *News Release: HHS launches campaign to prevent youth bullying. Washington, DC: DHHS.* Retrieved May 1, 2006, from http://www.hhs.gov/news/press/2004pres/20040301.html

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51,* 162-170.

Dodge, K. A. (1986). A social information-processing model of social competence in children. In M. Perlmutter, *Cognitive Perspectives on Children's Social and Behavioral Development: The Minnesota Symposia on Child Psychology* (pp. 77-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dodge, K. A. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74,* 374-393.

Dodge, K.A., Bates, J.E., & Pettit, G.S. (1990). Mechanisms in the cycle of violence. *Science, 250,* 1678-1683.

Dodge, K. A., & Coie, J. D. (1987). Social information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology, 53,* 1146-1158.

Dodge, K. A., Laird, R., Lochman, J. E., Zelli, A., & Conduct Problems Prevention Research Group (2002). Multidimensional latent-construct analysis of children's social information processing-patterns: Correlations with aggressive behavior problems. *Psychological Assessment, 14,* 60-73.

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., et al. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child Development, 74*, 374-393.

Dubow, E. F., Huesmann, L. R., & Boxer, P. (2003). Theoretical and methodological considerations in cross-generational research on parenting and child aggressive behavior. *Journal of Abnormal Child Psychology, 31*, 185-192.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology, 25,* 115-152.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology, 78,* 107-126.

D'Zurilla, T. J., & Nezu, A. M. (1990). Development and preliminary evaluation of the Social Problem-Solving Inventory. *Psychological Assessment, 2,* 156-163.

Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M. et al. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behaviors. *Child Development, 72,* 1112-1134.

Eisenberg, N., & Fabes, R. A. (1992). Emotion, regulation, and the development of social competence. In M. S. Clark (Ed.), *Emotion and social behavior: Vol. 14. Review of personality and social psychology* (pp. 119-150). Newbury Park, CA: Sage.

Eisenberg, N., & Fabes, R. A. (1999). Emotion, emotion-related regulation, and quality of socioemotional functioning. In L. B. Balter & C. Tamis-Lemonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 318-335). Philadelphia: Psychology Press.

Eisenberg, N., Sadovsky, A., Spinrad, T. L., Fabes, R. A., Losoya, S. H., Valiente, C., et al. (2005). The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity: Concurrent relations and prediction of change. *Developmental Psychology, 41,* 193-211.

Erdley, C. A., & Asher, S. R. (1996). Children's social goals and self-efficacy perceptions as influences on their responses to ambiguous provocation. *Child Development, 67*, 1329-1344.

Farrington, D. P., & Welsh, B. C. (1999). Delinquency prevention using family-based interventions. *Children & Society, 13*, 287-303.

Fast Track Project (1997). *Psychometric properties of the Social Health Profile.* Durham, NC: Duke University, Department of Psychology.

Fast Track Project (2003). *Teacher's Report Form for kindergarten/year 1: Fast Track Project technical report – November 26, 2003*. Retrieved November 6, 2004, from http://www.fasttrackproject.org/techrept/t/*TRF*/*TRF*1tech.pdf.

Federal Bureau of Investigation (2002). *Crime in the United States, 2001.* Washington, DC: Federal Bureau of Investigation, Table 41.

Fontaine, R. G., Burks, V. S., & Dodge, K. A. (2002). Response decision processes and externalizing behavior problems in adolescents. *Development and Psychopathology, 14,* 107-122.

Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots*. Ontario, CA: Sage Publications, Inc

Fraser, M. W. (1996). Cognitive problem-solving and aggressive behavior among children. *Families in Society, 77*, 19-32.

Fraser, M. W. (2004). *Risk and Resilience in Childhood: An Ecological Perspective* (2nd ed.). Washington, DC: NASW Press.

Fraser, M. W., Day, S. H., Galinsky, M. J., Hodges, V. G., & Smokowski, P. R. (2004). Conduct problems and peer rejection in childhood: A randomized trial of the Making Choices and Strong Families programs. *Research on Social Work Practice, 14,* 313–324.

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. (2005). Social information-processing skills training to promote social competence and prevent aggressive behavior in the third grade. *Journal of Consulting and Clinical Psychology, 73*, 6, 1045–1055.

Fraser, M. W., Nash, J. K., Galinsky, M. J., & Darwin, K. M. (2000). *Making Choices*: Social problem-solving skills for children. Washington, DC: NASW Press.

Fraser, M. W., Rose, R. A., Terzian, M. A., & Guo, S. (2004, October). Propensity Score Matching using STATA/PSMATCH2. Paper presented to the Applied Statistics Working Group at the University of North Carolina at Chapel Hill, Chapel Hill, NC.

Galinsky, M. J., & Terzian, M. A. (2006). The art of group work practice with manualized curricula. *Social Work with Groups, 29*, 11-26.

Gillespie, D. F. & Streeter, C. L. (1994). Fitting regression models to research questions for analyzing change in non-experimental research. *Social Work Research, 18,* 4, 239-245.

Gottfedson, D. C. (1998). School-based crime prevention. In L.W. Sherman, D. C. Gottfredson, D. L. MacKenzie, J. Eck, P. Reuter, and S. Bushway, *Preventing crime: what works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: US Department of Justice, National Criminal Justice Reference Service. Retrieved October 1, 2006, from http://www.ncjrs.gov/works/index.htm.

Gottfredson , G. D & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Educational & Psychological Consultation,* 12, 313-344.

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Greenberg et al. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466–474

Greenberg, M. T., Kusché, C. A., Cook, E. T., & Quamma, J. P. (1995). Promoting emotional competence in school-aged children - the effects of the *PATHS* Curriculum. *Development and Psychopathology, 7*, 117-136.

Greenberg, M.T.,& Kusché, C.A. (1998). Preventive intervention for school-aged deaf children: The *PATHS* curriculum. *Journal of Deaf Studies and Deaf Education, 3*, 49-63.

Greenhill, L. L., & Malcolm, J. (2000). Diagnostic measures in child disorders. In J. Rush and H. Pincus (Eds.), *Handbook of psychiatric measures*. Washington, DC: American Psychiatric Association.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85,* 348 –362.

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

Hanish, L. D., & Guerra, N. G. (2000). Predictors of peer victimization among urban youth. *Social Development, 9,* 521-543.

Hanish, L. D., & Guerra, N. G. (2002). A longitudinal analysis of patterns of adjustment following peer victimization. *Development and Psychopathology, 14,* 69-89.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1996). Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13416-13420.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from Evaluating a job training programme. *The Review of Economic Studies*, 64, 605-654.

Heckman, J. J., & Robb, R. (1985). Alternative methods for evaluating the impact of treatment on the outcomes, In J. Heckman and B. Singer (Eds.), *Longitudinal analysis of labor market data.* Cambridge: Cambridge University Press.

Hox, J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Lawrence Erlbaum.

Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *BehavioralAssessment, 11*, 459-467.

Hubbard, J. A. (2001). Emotion expression processes in children's peer interaction: The role of peer rejection, aggression, and gender. *Child Development, 72,* 1426-1438.

Hubbard, J. A., Dodge, K. A., Cillessen, A. H. N., Coie, J. D., & Schwartz, D. (2001). The dyadic nature of social information processing in boys' reactive and proactive aggression. *Journal of Personality and Social Psychology, 80,* 268-280.

Hudley, C. & Graham, S. (1993). An attributional intervention to reduce peer-directed aggression among African-American boys. *Child Development, 64*, 124-138.

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior*, *14*, 13-24

Huesmann, L. R. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for social policy* (pp. 73-109). New York: Academic Press.Howe, G. W., Reiss, D., & Yuh, J. (2002). Can prevention trials test theories of etiology? *Development and Psychopathology, 14,* 673-694.

Huesmann, L. R., Eron, L. D., Guerra, N. G., & Crawshaw, V. B. (1994). Measuring children's aggression with teachers' predictions of peer nominations. *Psychological Assessment, 6,* 329-336.

Hughes, J. N., Meehan, B. T., & Cavell, T. A. (2004). Development and validation of a gender-balanced measure of aggression-relevant social cognition. *Journal of Clinical Child and Adolescent Psychology, 33*, 292-302.

Izard, C.E. (2002). Translating emotion theory and research into preventative interventions. *Psychological Bulletin, 128*, 796-824.

Johnson, L. C., Beaton, R., & Murphy, S.A. (2005). First record: A methodological approach to counter sampling bias. *Psychological Reports, 95*, 391-392.

Kam, C. M., Greenberg, M. T., & Kusché, C. A. (2004). Sustained effects of the *PATHS* curriculum on the social and psychological adjustment of children in special education. *Journal of Emotional and Behavioral Disorders, 12*, 66-78.

Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the *PATHS* curriculum. *Prevention Science, 4,* 55-63.

Kay S. R., Wolkenfield F., & Murrill L. M. (1998) Profiles of aggression among psychiatric patients. *Journal of Nervous and Mental Disease*, 176, 547–557.

Kazdin, A. E. & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44,* 1116–1129.

Kazdin, A. E., Siegel, T. C., & Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology, 60,* 733-747.

Kellam, S. B., Branch, J. D., Agrawal, K. C., & Ensminger, M. E. (1975). *Mental health and going to school: The Woodlawn program of assessment, early intervention and evaluation.* Chicago: University of Chicago Press.

Kellam, S. G., & Rebok, G. W. (1992). Building developmental and etiological theory through epidemiologically based preventive intervention trials. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 162-195). New York, NY: Guilford Press.

Kempes, M., Matthys, W., de Vries, H., & van Engelend, H. (2005). Reactive and proactive aggression in children: A review of theory, findingds, and the relevance for child and adolescent psychiatry (2005). *European Child and Adolescent Psychiatry, 14*, 11-19.

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Annals of Family Medicine, 2,* 204-208.

Kline, R. B. (1998). *The principles and practice of structural equation modeling*, New York: Guilford Press.

Kochenderfer-Ladd, B. (2003). Identification of aggressive and asocial victims and the stability of their peer victimization. *Merrill-Palmer Quarterly, 49*, 401-425.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*, 418-444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277.

Lahey, B. B., Schwab-Stone, M., Goodman, S. H., Waldman, I. D., Canino, G., Rathouz, P. J. et al. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology, 109,* 488-503.

Landau, S., & Moore, L. A. (1991). Social skill deficits in children with attention-deficit hyperactivity disorder. *School Psychology Review, 20,* 235–251.

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71*, 107-118.

Lengua, L. J. (2003). Associations among emotionality, self-regulation, adjustment problems, and positive adjustment in middle childhood. *Journal of Applied Developmental Psychology, 24*, 595-618.

Leslie, L. K., Gordon, J. N., Lambros, K., Premji, K., Peoples, J., & Gist, K. (2005). Addressing the developmental and mental health needs of young children in foster care (Special Article). *Journal of Developmental & Behavioral Pediatrics, 26*, 140-152.

Limber, S. P., & Nation, M. M. (1998). Bullying among children and youth. In J. L. Arnette & M. C. Walsleben (Eds.) *Juvenile Justice Bulletin: Combating fear and restoring safety in schools, April issue.* Retrieved June 1, 2005, from http://ojjdp.ncjrs.org/jjbulletin/ 9804/bullying2.html.

Lindsay, E. W., & Kurtz, D. P. (1987). Evaluation of a school-juvenile court team approach to delinquency prevention. *Children & Youth Services Review, 9,*101-115.

Lipsey, M. W. (1988). Practice and malpractice in evaluation research. *Evaluation Practice, 9,* 5-24.

Lipsey, M. W. & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D.P. Farrington (Eds.), *Serious & violent juvenile offenders: Risk factors and successful interventions* (pp. 86-105). Thousand Oaks, CA: Sage.

Little, T. D., Jones, S. M, & Henrich, C. C. (2003). Disentangling the "whys" from the "whats" of aggressive behavior. *International Journal of Behavioral Development, 27*, 122-133.

Little, T. D**.,** Lindenberger, U. & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192-211.

Lochman, J. E. (1992). Cognitive-behavioral intervention with aggressive boys: Three year follow-up and preventive effects. *Journal of Consulting & Clinical Psychology, 60, 426*-432.

Lochman, J. E., Burch, P. R., Curry, J. F. & Lampron, L. B. (1984). Treatment and generalization effects of cognitive-behavioral and goal-setting interventions with aggressive boys. *Journal of Consulting and Clinical Psychology, 52*, 915-916.

Lochman, J. E., & Curry, J. F. (1986). Effects of social problem-solving training and self-instruction training with aggressive boys. *Journal of Clinical Child Psychology, 15*, 159-164.

Lochman, J. E., & Dodge, K. A. (1994). Social-cognitive processes of severely violent, moderately aggressive, and nonaggressive boys, *Journal of Consulting and Clinical Psychology, 62,* 366-374.

Lochman, J. E., & Dodge, K. A. (1998). Distorted perceptions in dyadic interactions of aggressive and nonaggressive boys: Effects of prior expectations, context, and boys' age. *Development and Psychopathology, 10,* 495-512.

Lochman, J. E., & Lampron, L. B. (1988). Cognitive behavioral interventions for aggressive boys: Seven months follow-up effects. *Journal of Child and Adolescent Psychotherapy, 5,* 15-23.

Lochman, J. E., Lampron, L. B., Gemmer, T. C., Harris, S. R., & Wyckoff, G. M. (1989). Teacher consultation and cognitive-behavioral intervention with aggressive boys. *Psychology in the Schools, 26*, 179-188.

Lochman, J. E., & Lenhart, L. A. (1993). Anger Coping Intervention for Aggressive-Children - Conceptual Models and Outcome Effects. *Clinical Psychology Review, 13,* 785-805.

Lochman, J. E., & Salekin, R. T. (2003). Behaviorally oriented interventions for children with aggressive behavior and/or conduct problems. *Behavior Therapy, 34*, 413-419.

Lochman, J. E., & Wells, K. C. (2002). Contextual social-cognitive mediators and child outcome: A test of the theoretical model in the Coping Power program. *Development and Psychopathology, 14,* 945-967.

Lochman, J. E., & Wells, K. C. (2003). Effectiveness of the coping power program and of classroom interventions with aggressive children: Outcomes at a 1-year follow-up. *Behavior Therapy, 34*, 493-515.

Lochman, J. E., & Wells, K. C. (2004). The coping power program for preadolescent aggressive boys and their parents: Outcome effects at the 1-year follow-up. *Journal of Consulting & Clinical Psychology, 72,* 571-578.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). *Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence*. Mahwah, NJ: Lawrence Erlbaum Publishers.

Losel, F., & Beelman, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized interventions. *The Annals of the American Academy of Political and Social Science, 587,* 84-109.

Macgowan, M. J., Nash, J. K., & Fraser, M. W. (2002). The Carolina Child Checklist of risk and protective factors for aggression. *Research on Social Work Practice, 12*, 253-276.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.

MacKinnon, D. P., Taborga, M. P., Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence, 68,* S69-S83.

Maedgen, J. W., & Carlson, C. L. (2000). Social functioning and emotional regulation in the attention deficit hyperactivity disorder subtypes. *Journal of Clinical Child Psychology, 29,* 30–42.

Making Choices Project (2004). *Inter-coder reliability of the Skill Level Activity encoding measure using Cohen's Kappa: Technical report*. Chapel Hill, NC: The University of North Carolina at Chapel Hill, School of Social Work.

Matthys, W., Cuperus, J. M., & Van Engeland, M. (1999). Deficient social problem-solving in boys with ODD/CD, with ADHD, and with both disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 311-321.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

McDonald, L., Billingham, S., Conrad, T., Morgan, A., O, N., & Payton, E. (1997). Families and schools together (FAST). Integrating community development with clinical strategies. *Families in Society-the Journal of Contemporary Human Services, 78,* 140-155.

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

McMahon, S. D., Washburn, J. J., Felix, E. D., Yakin, J., & Childrey, G. (2000). Violence prevention: Program effects on urban preschool and kindergarten children. *Applied and Preventive Psychology, 9*, 271-281.

Merrell, K. W., Sanders, D. E., & Popinga, M. (1993). Teacher ratings of social behavior as a predictor of special education status: Discriminant validity of the School Social Behavior Scales. *Journal of Psychoeducational Assessment, 11,* 220-231.

Mihalic, S., Irwin, K., Fagan, A., Ballard, D., & Elliot, D. (2004, July). Successful program implementation: lessons from Blueprints. *Juvenile Justice Bulletin*. Washington, DC: Office of Juvenile Justice Delinquency Prevention.

Milich, R., & Dodge, K. A. (1984). Social information processing patterns in child psychiatric populations. *Journal of Abnormal Child Psychology, 12,* 471-490.

Miller-Johnson, S., Coie, J,D., Maumary-Gremaud, A; & Bierman, K. (2002). Peer rejection and aggression and early starter models of conduct disorder. *Journal of Abnormal Child Psychology*, *30,* 217-231.

Miller-Johnson, S., Winn, D. M. C., Coie, J. D., Malone, P. S., & Lochman, J. (2004). Risk factors for adolescent pregnancy reports among African American males. *Journal of Research on Adolescence, 14*, 471-495.

Moffitt, T. E. (2003). Life-course-persistent and adolescence-limited antisocial behavior: A 10-year research review and a research agenda. In B. B. Lahey, T. E. Moffitt, & A. Caspi, *Causes of conduct disorder and juvenile delinquency* (pp. 49-75). New York: Guilford Press.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescent-limited antisocial pathways among males and females. *Development and Psychopathology, 13,* 355-375.

Morgan, G. A., Gliner, J. A., & Harmon, R. J. (2000). Internal validity. *Journal of the American Academy of Child & Adolescent Psychiatry. 39,* 529-531.

Mplus (n.d.). *Chi-square difference testing using the Satorra-Bentler scaled chi-square.* Retrieved October 1, 2006, from http://www.statmodel.com/chidiff.shtml

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

Musher-Eizenman, D. R., Boxer, P., Danner, S., Dubow, E. F., Goldstein, S. E., & Heretick, D. M. L. (2004). Social-cognitive mediators of the relation of environmental and emotion regulation factors to children's aggression. *Aggressive Behavior, 30*, 389-408.

Muthén, B. O. (2006). Traditional latent variable modeling using Mplus. Mplus Workshop presented on October 19, 2006. Baltimore, MD: Johns Hopkins University.

Muthén, L. K. (1999). Mplus discussion board: Multilevel data/complex sample. Retrieved October 1, 2006, from http://www.statmodel.com/discussion/messages/12/18.html?SaturdayApril820000848am

Muthén, L. K., & Muthén, B.0. (1998-2006). Mplus user's guide (4th ed.). Los Angeles, CA: Muthén & Muthén.

Nagin, D. S., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70,* 1181-1196.

Nansel, T.R., Overpeck, M.D., Pilla, R.S., Ruan, W.J., Simons-Morton, B. & Scheidt, P.C. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association*, 285, 2094-2100.

Nash, J. K., Fraser, M. W., Galinsky, M., & Kupper, L. (2003). Early development and pilot-testing of a problem-solving skills-training program for children. *Research on Social Work Practice, 13*, 432 – 450.

Nash, J. K., Kupper, L., & Fraser, M. W. (2004). Using multilevel statistical models in social work intervention research. *Journal of Social Service Research, 30,* 35-54.

Nelson, D. A., & Crick, N. R. (1999). Rose-colored glasses: Examining the social information-processing of prosocial young adolescents. *Journal of Early Adolescence, 19*, 17-38.

NICHD Early Child Care Research Network. (2004). Trajectories of physical aggression from toddlerhood to middle childhood: Predictors, correlates, and outcome. W. F. Overton (Ed.), *Monographs of the Society for Research in Child Development* (Serial no. 278, vol. 69, no. 4). Boston, MA: Blackwell Publishing.

Nunnaly, J. (1978). Psychometric theory. New York: McGraw-Hill.

O'Donnell, J., Hawkins, J. D., Catalano, R. F., Abbott, R. D., & Day, E. (1995). Preventing school failure, drug use, and delinquency among low-income children: Long-term intervention in elementary schools. *American Journal of Orthopsychiatry, 65*, 87-100.

Office of Special Education and Rehabilitative Services (2000). *A guide to the Individualized Education Program*. Retrieved May 1, 2006, from Department of Education online via: http://www.ed.gov/parents/needs/speced/iepguide/iepguide.pdf

Orobio de Castro, B., Bosch, J. D., Veerman, J. W., & Koops, W. (2003). The effects of emotion regulation, attribution, and delay prompts, on aggressive boys' social problem solving. *Cognitive Therapy and Research, 27,* 153–166.

Orobio de Castro, B., Merk, W., Koops, W., Veerman, J. W., & Bosch, J. D. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology, 34*, 105-116.

Orobio de Castro, B., Veerman, J. W., Koops, W., Bosch, J. D., Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development, 73,* 916-934.

Orpinas, P., Parcel, G. S., McAlister, A., & Frankowski, R. (1995). Violence prevention in middle schools: A pilot evaluation. *Journal of Adolescent Health, 17,* 360-371.

Ozer, E. J. (2006). Contextual effects in school-based violence prevention programs: A conceptual framework and empirical review. *The Journal of Primary Prevention, 27*, 317-342.

Patterson, G. R., Forgatch, M. S., Yoerger, K. L., & Stoolmiller, M. (1998). Variables that initiate and maintain an early-onset trajectory for juvenile offending. *Development and Psychopathology, 10*, 531-547.

Patterson, G. R., & Yoerger, K. L. (1993). Developmental models for delinquent behavior. In S. Hodgins, *Mental disorder and crime* (pp. 140-172). Thousand Oaks: Sage.

Pelham, W.E., Massetti, G.M., & Waschbusch, D.A. (2005, April). Effects of a school-wide SACD intervention on children with disruptive behavior problems. In T.M. Haegerich & L.E. Reese (Chairs), *The Multi-Site Social and Character Development Research Program: Innovations in theory, methodology, and statistical techniques*. Symposium conducted at the biannual meeting of the Society for Research in Child Development, Atlanta, Georgia.

Petrosino, A. (2000). Mediators and moderators in the evaluation of programs for children. *Evaluation Review, 24,* 47-72.

Policy Design Team. (1994). *Mental health needs of youth in Virginia's juvenile detention centers.* Department of Criminal Justice Services: Richmond, VA.

Pope, A. W., Bierman, K. L., & Mumma, G. H. (1989). Relations between hyperactive and aggressive behaviors and peer relations at three elementary grade levels, *Journal of abnormal child psychology*, *17,* 253-267.

Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression: evidence of a two-factor model. *Psychological Assessment, 12,* 115-122.

Poulin, F., Dishion, T. J., & Burraston, B. (2001). Three-year iatrogenic effects associated with aggregating high-risk adolescents in cognitive–behavioral preventive interventions. *Applied Developmental Science, 5,* 214–224.

Preacher, K. J., & Hayes, A. F. (in review). *Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models*.

Prinstein, M. J., & Cilessen, A. H. (2003). Forms and functions of adolescent peer aggression associated with high levels of peer status. *Merrill-Palmer Quarterly, 49,* 310–342.

Prinstein, M. J., & La Greca, A. M. (2004). Childhood peer rejection and aggression as predictors of adolescent girls' externalizing and health risk behaviors: A 6-year longitudinal study. *Journal of Consulting and Clinical Psychology, 72*, 103-112.

Putallaz, M , & Bierman, K. L. (2004). *Aggression, antisocial behavior, and violence among girls: A developmental perspective*. New York: Guilford Press.

Quiggle, N. L., Garber, J., Panak, W. F., & Dodge, K. A. (1992). Social information processing in aggressive and depressed children. *Child Development, 63*, 1305-1320.

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

Raudenbush, S. W., & Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods* 6, 387-401.

Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2005). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Retrieved August 14, 2005, from http://sitemaker.umich.edu/group-based/optimal_design_software

Reid, J. B., Eddy, J. M., Fetrow, R., & Stoolmiller, M. (1999). Description and immediate impacts of a preventive intervention for conduct problems. *American Journal of Community Psychology*, 27, 483-517.

Rhodes, W., Pelissier, B., Gaes, G. Saylor, W., Camp, S., & Wallace, S. (2001). Alternative solutions to the problem of selection bias in an analysis of federal residential drug treatment programs. *Evaluation Review, 25,* 331-369.

Richard, B. A., & Dodge, K. A. (1982). Social maladjustment and problem-solving in school-aged children. *Journal of Consulting and Clinical Psychology, 50,* 226-233.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes. S*an Diego, CA: Academic Press.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Rubin, A., & Babbie, E. (1997). *Research methods for social work (3rd ed.).* Pacific Grove, CA: Brooks/Cole.

Rutter, M. (1997). Nature-nurture integration: The example of antisocial behavior. *American Psychologist, 52*, 390-398.

Rydell, A. M., Berlin, L., Bohlin, G. (2003). Emotionality, emotion regulation, and adaptation among 5- to 8-year-old children. *Emotion, 3,* 30-47.

Sameroff, A. J. (1985). Environmental factors in the early screening of children at risk. In W. Frankenburg, W.R. Emde, and J. Sullivan (Eds.), *Early identification of children at risk: an international perspective* (pp. 21-44). New York: Plenum Press.

SAS, Inc. (2000). *SAS/SYSTAT User's Guide: Version 8, Volumes 1, 2, & 3*. Cary, NC: Author.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp.233-247). London: Kluwer Academic Publishers.

Shaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39,* 1020-1035.

Schultz, D., Izard, C.E., & Bear, G.G. (2004). Children's emotion processing: Relations to emotionality and aggression. *Development and Psychopathology, 16,* 371-387.

Schwartz, D., Dodge, K.A., Coie, J. D., Hubbard, J. A., Cillessen, A. H., Lemerise, E. A. et al. (1998). Social-cognitive and behavioral correlates of aggression and victimization in boys' play groups. *Journal of Abnormal Child Psychology, 26,* 431-440.

Schwartz, D., McFayden-Ketchum, S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (1999). Early behavior problems as a predictor of later peer group victimization: Moderators and mediators in the pathways of social risk. *Journal of Abnormal Child Psychology, 27*, 191-201.

Seipel, S. J., & Apigian, C. H. (2005). Perfectionism in students: Implications in the instruction of statistics. *Journal of Statistics Education*, *13.* Retrieved May 1, 2006, from: http://www.amstat.org/publications/jse/v13n2/seipel.html

Shackman, G. (2001). *Sample size and design effect. New York State Department of Health* (NYS DOH). Presented on March 24, 2001 at Albany Chapter of American Statistical Association. Retrieved October 1, 2006, from http://faculty.smu.edu/slstokes/stat3370/deff.pdf.

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

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445.

Shure, M. B., & Spivack, G. (1980). Interpersonal problem solving as a mediator of behavioral adjustment in preschool and kindergarten children. *Journal of Applied Developmental Psychology, 1*, 29-44.

Smithmyer, C. M., Hubbard, J. A., Simons, R. F. (2000). Proactive and reactive aggression in delinquent adolescents: relations to aggression outcome expectancies. *Journal of Clinical Child Psychology, 29,* 86-93.

Smokowski, P. R., Fraser, M. W., Day, S. H., Galinsky, M. J., & Bacallao, M. L. (2004). School-based skills training to prevent aggressive behavior and peer rejection in childhood: Evaluating the *Making Choices* program. *The Journal of Primary Prevention, 25,* 233-251.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis*. London: Sage.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290-312). Washington, DC: American Sociological Association.

Spivack, G., & Shure, M. B. (1985). ICPS and Beyond - Centripetal and Centrifugal Forces. *American Journal of Community Psychology, 13,* 226-243.

Spoth, R., Redmond, C., & Shin, C. (1998). Direct and indirect latent-variable parenting outcomes of two universal family-focused preventive interventions: Extending a public health-oriented research base. *Journal of Consulting and Clinical Psychology, 66,* 385-399.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system (2nd Ed.).* Cary, NC: SAS Institute, Inc.

Stormshak, E. A., Bierman, K. L., & The Conduct Problems Prevention Research Group (1998). The implications of different developmental patterns of disruptive behavior problems for school adjustment. *Development and Psychopathology, 10,* 451-467.

Tapper, K., & Boulton, M. J. (2004). Sex differences in levels of physical, verbal, and indirect aggression amongst primary school children and their associations with beliefs about aggression. *Aggressive Behavior, 30,* 123-145.

Taub, J. (2002). Evaluation of the *Second Step* violence prevention program at a rural elementary school. *School Psychology Review, 31,* 186-200.

Tein, J., Sandler, I. N., MacKinnon, D. P., & Wolchik, S. A. (2004). How did it work? Who did it work for? Mediation in the context of a moderated prevention effect for children of divorce. *Journal of Consulting and Clinical Psychology, 72,* 617–624

Terzian, M. A., & Fraser, M. W. (2005). Preventing aggressive behavior and drug use in elementary school: Six family-oriented programs. *Aggression and Violent Behavior, 10,* 407-435.

Tolan, P. (2001). Youth violence and its prevention in the United States: An overview of current knowledge. *Injury Control and Safety Promotion, 8*, 1-12.

Trachtenberg, S., & Viken, R. J. (1994). Aggressive boys in the classroom: Biased attributions or shared perceptions? *Child Development, 65,* 829-835.

Tremblay, R. E., Masse, B., Perron, D., Leblanc, M., Schwartzman, A. E., & Ledingham, J. E. (1992). Early disruptive behavior, poor school achievement, delinquent behavior, and delinquent personality: Longitudinal analyses. *Journal of Consulting and Clinical Psychology, 560,* 64-72.

Tremblay, R. E., Pagani-Kurtz L., Masse, L. C., Vitaro, F., & Pihl, R. O. (1995). A bimodal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. *Journal of Consulting and Clinical Psychology, 63,* 560-568.

Underwood, M. K., & Coie, J. D. (2004). Future directions and priorities for prevention and intervention. In M. Putallaz & K. L. Bierman, *Aggression, antisocial behavior, and violence among girls* (pp.289-301). New York: Guilford Press.

Van Schoiack-Edstrom, L., Frey, K. S., & Beland, K. (2002). Changing adolescents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. *School Psychology Review, 31,* 201-216.

Vermeiren, R., Jespers, I., & Moffitt, T. (2006). Mental health problems in juvenile justice populations. *Child and Adolescent Psychiatric Clinics of North America. 15*, 333-351.Vernberg, E. M., Jacobs, A. K., & Hershberger, S. L. (1999). Peer victimization and attitudes about violence during early adolescence. *Journal of Clinical Child Psychology, 28,* 386-395.

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development, 30,* 12-19.

Vitaro, F., Brendgen, M., Pagani, L., Tremblay, R. E., & McDuff, P. (1999). Disruptive behavior, peer association, and conduct disorder: Testing the developmental links through early intervention. *Development and Psychopathology, 11*, 287-304.

Vitaro, F., Gendreau, P. L., Tremblay, R. E., & Oligny, P. (1998). Reactive and proactive aggression differentially predict later conduct problems. *Journal of Child Psychology and Psychiatry, 39,* 377-385

Walden, T. A., & Smith, M. C. (1997). Emotion regulation. *Motivation and Emotion, 21*, 7-25.

Waldman, I. D. (1996). Aggressive boys' hostile perceptual and response biases: The role of attention and impulsivity. *Child Development, 67*, 1015-1033.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M.J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4,* 194-209.

Wall, A. E. (2005). Problem behaviors in maltreated children and youth: Influential child, peer, and caregiver characteristics, 1999-2000. Doctoral Dissertation.

Wasserman, G. A., Ko, S. J., & McReynolds, L. S. (2004). Assessing the mental health status of youth in juvenile justice settings. *Juvenile Justice Bulletin, August*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Werthamer-Larsson, L., Kellam, S. G., & Wheeler, L. (1991). Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19,* 585-602.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71,* 136-149.

Woodward, L. J., & Fergusson, D. M. (1999). Early conduct problems and the risk of teenage pregnancy in girls. *Development and Psychopathology, 11,* 127-141.

Zelli, A., Dodge, K. A., Lochman, J. E., Laird, R. D., & Conduct Problems Prevention Research Group (1999). The distinction between beliefs legitimizing aggression and deviant processing of social cues: Testing measurement validity and the hypothesis that biased processing mediates the effects of beliefs on aggression. *Journal of Personality and Social Psychology, 77,* 150-166.

Zyzanski, S. J., Flocke, S. A., & Dickinson, L. M. (2004). On the nature and analysis of clustered data. *Annals of Family Medicine, 2,* 199-200.