SOCIAL INFORMATION PROCESSING AND AGGRESSIVE BEHAVIOR IN CHILDHOOD: THEORY AND PRACTICE

Jilan Li

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
2012

Approved by:

Chair: Mark W. Fraser, Ph.D.

Susan T. Ennett, Ph.D.

Maeda J. Galinsky, Ph.D.

Joelle D. Powers, Ph.D.

Kathleen A. Rounds, Ph.D.

Christopher A. Wiesen, Ph.D.

©2012
Jilan Li
ALL RIGHTS RESERVED

ABSTRACT

JILAN LI: Social Information Processing and Aggressive Behavior in Childhood: Theory and Practice
(Under the direction of Dr. Mark W. Fraser)

Social-emotional skills training is ubiquitous in American public schools; however, the effectiveness of these programs has not been well-established. Small effect sizes plus mixed and contradictory findings raise the importance of refining existing programs by incorporating new knowledge in social cognitive and behavioral sciences and investigating factors that contribute to discrepancies across evaluation findings. This three-paper dissertation is an effort to address these issues.

The first paper reviews an important theoretical advance in social cognitive research: the social information processing (SIP) theory. The paper develops a general framework for applying SIP theory to social-emotional skills training, and reviews issues in applying SIP to practice. The paper distinguishes SIP-based interventions from traditional social problem-solving (SPS) interventions. Several methodological issues in conducting SIP intervention research are discussed.

The second and the third papers investigate one implementation factor—the length of treatment exposure or *dosage*—to help explain the contradictory findings from evaluation studies of social-emotional skills training programs. Investigating the effects of varying dosage (i.e., *dosage analysis*) is an important but critically understudied area of social intervention research. Dosage analysis requires advanced statistical techniques to balance multiple dosage groups and estimate valid effects by treatment exposure level. The second paper reviews a recent development in the family of propensity score-based methods—generalized propensity score-based (GPS) methods—with potential utility for balancing multiple dosage groups. In addition to discussing GPS application principles, this paper demonstrates the use of one GPS method with a continuous treatment variable.

The third paper investigates dosage effects of a SIP-based social-emotional skills training program, the *Making Choices* program. The analysis uses the GPS method with a continuous treatment variable. Data were drawn from a national evaluation study of *Making Choices*. Dosage effects were evaluated for eight key outcomes at the end of Grade 3 and Grade 4 years. Findings indicate dosage effects on social competence and emotional regulation at the end of Grade 3. No effects were observed at the end of Grade 4. Further, findings suggest characteristics of the quality of implementation (e.g., level of student engagement, teacher-student relationship) are important areas for future investigation.

For Mom and Dad

ACKNOWLEDGMENTS

Seven and a half years is a significant portion of a lifetime and a long time to spend on a single journey— the MSW/PhD program. I am grateful to have a formal opportunity to express my deep appreciation and gratitude to those who have helped me complete this journey. These individuals not only offered me much-needed support but also kept me mindful of the beauty of life, which helped keep my spirits up as I persevered to finish the program.

I thank Mark Fraser, my dissertation chair and supervisor, for demonstrating the beauty of research through his way of conducting research and his way of sharing his time, knowledge, and wisdom with unbounded generosity. He never pushed me, but he kept me motivated to do more research.

I thank each of my dissertation committee members, Susan Ennett, Shenyang Guo, Maeda Galinsky, Kathleen Rounds, Joelle Powers, and Christopher Weisen, for sharing their wealth of expertise and for offering me their encouragement and compassion. A special thank-you goes to Christopher Weisen for his help and support that was critical in the data analysis process.

I also thank the team members of the Social and Character Development project, Mark Fraser, Shenyang Guo, Steve Day, Alan Ellis, Traci Wike, and Rod Rose, for making our team meetings such thought-provoking and exciting learning experiences. A special thank-you goes to Steve Day for his generosity in sharing his time and encyclopedic knowledge of intervention and data collection procedures.

The journey of doctoral study has always been intertwined with the complexity of life. I often needed advice not only to achieve academic success but also to handle the challenges and difficulties presented by life. For providing caring heart, a listening ear, and sound advice, I thank Kathleen Rounds, Mark Fraser, Maeda Galinsky, and Natasha Bowen; thank-you for being so open and available. This journey has been possible because I felt free to talk to each of you whenever I needed.

Scientific writing is challenging, especially when writing in a second language. I am deeply grateful to Diane Wyant for her excellent and generous editorial help and for sharing her humor that often lightened up a gloomy day.

I am fortunate to have a family that is bound closely with love. My journey through the doctoral program has truly been a team effort of my whole family. My mother, my sisters, my brother, and my brother-in-law came from China to help take care of my daughter; their sacrifice allowed me to leave for school before my daughter woke up and to return late in the evening to find a happy baby and delicious meals on the table. Special thanks go to my husband Binggang, who did what I needed from him during my last big push to finish. I also thank my daughter April, who contributed to my success in her own way. She never cried when I told her, "Mom needs to go to school." She never complained when I could not spend weekend time with her even when she was sick. Thank you, April. I promise I will spend more time with you.

Thank you to all the people who made the long journey not just possible but also beautiful!

TABLE OF CONTENTS

LIST OF TABLES

Table

INTRODUCTION

SOCIAL INFORMATION PROCESSING AND AGGRESSIVE BEHAVIOR IN CHILDHOOD: THEORY AND PRACTICE

Social-emotional skills training programs are ubiquitous in American public schools. The wide implementation of these programs is based on two findings from developmental research: early aggressive behavior is associated with poor developmental outcomes (e.g., Dodge & Pettit, 2003; Fraser, 1996; Moffitt & Caspi, 2001; Prinstein & La Greca, 2004; Odgers et al., 2008) and social-emotional skills deficits appear to mediate early aggressive behavior and a variety of later conduct problems (e.g., Bandura, 1989; Dodge, 1980, 2006; Huesmann, 1988; Lengua, 2003; Zins, Weissberg, Wang, & Walberg, 2004).

Systematic reviews of evaluation studies of universal school-based social-emotional skills training have suggested that the majority of these programs were effective (Farrington & Welsh, 2003; Hahn et al., 2007; Payton et al., 2008; Wilson & Lipsey, 2003, 2006, 2007). However, the effect sizes were moderate, at most. Moreover, a number of studies have reported mixed or even negative effects (e.g., Conduct Problems Prevention Research Group [CPPRG], 1999; Flannery et al., 2003; Grossman et al., 1997; Malti, Ribeaud, & Eisner, 2011; Merrell, Gueldner, Ross, & Isava, 2008; Park-Higgerson, Perumean-Chaney, Bartolucci, Grimley, & Singh, 2008; Multisite Violence Prevention Project, 2009). Even those programs shown to be effective in previous rigorous evaluation studies often showed substantially reduced or no effects when evaluated independent of the program developers (Eisner, 2009; Malti, Ribeaud, & Eisner, 2011).

The lack of strong evidence supporting the effectiveness of social-emotional skills training suggests the need to improve these programs by incorporating new knowledge of children's cognition, emotion, and behavior. Over the past two decades, the understanding of the cognitive mechanisms underlying aggressive behavior among children has been substantively advanced through the development of SIP theory. However, the translation of the SIP perspective to practice is still in a formative stage with few applications to real-world settings. Notably, only a few school-based programs have explicitly used the SIP model to guide their curriculum design (e.g., Fraser et al., 2005; Meyer & Farrell, 1998). Moreover, although some researchers have used SIP theory in tandem with other theories (e.g., social learning theory, cognitive scripts) to design interventions, few have acknowledged the unique contribution of SIP theory.

An issue of greater concern is the frequent confusion of the traditional social problem-solving (SPS) approach (D'Zurilla & Goldfried, 1971; Shure & Spivack, 1988; Spivack & Shure, 1974) with the SIP model (e.g., Frey, Hirschstein, & Guzzo, 2000; Wilson & Lipsey, 2006), even though the two models outline distinct theoretical approaches. The blending of a SIP intervention with traditional SPS interventions obscures the translation of advances in theory to practice. Before SIP theory can be adequately and appropriately incorporated in the design of interventions, it is critical that the field first develop a clear understanding of the SIP theory and clearly distinguish between the SIP approach and a traditional SPS approach.

Another important issue in social-emotional skills training is to understand the discrepancy between findings across evaluation studies. The observed effects of a program are produced through a complex process on which a variety of factors can impinge. An important factor is the length of participants' actual exposure to the training, that is, the program *dosage*. The dosage received by each participant might vary widely for many reasons (e.g., only half of the treatment was delivered, some participants missed half of the training classes). Variation in intervention dosage would result in different treatment effects. Therefore, it is important that program evaluation goes beyond the estimation of overall treatment effects to further examine if and in what ways responses vary by the length of treatment exposure (i.e., dosage). Dosage analysis represents an emerging line of inquiry with the potential to help untangle core factors affecting treatment effects.

Assessing Effects by Level of Intervention Exposure: Dosage Analyses

Assessing dosage effects is challenging because groups with varying program dosages are often not purposefully formed through randomization, but are formed as a result of varying program implementation; therefore, the groups are not directly comparable and statistical measures have to be taken to account for overt selection bias (i.e., bias due to differences in observed covariates, Rosenbaum, 1991). Moreover, dosage analyses often involve comparing more than two groups. Balancing multiple groups simultaneously is particularly challenging and often requires advanced statistical methods rather than relying on conventional regression analyses or matching techniques generally used for two-group comparisons.

The recent development of generalized propensity score (GPS) methods has provided researchers with a viable means to balance multiple groups simultaneously. However, discussions of these methods have been largely confined to statisticians and economists (e.g., Hirano & Imbens, 2004; Imai & Van Dyk, 2004; Imbens, 2000; Joffe & Rosenbaum, 1999). Introducing these methods to social science researchers is an important step in addressing issues related to dosage analysis and, in turn, addressing issues in social- emotional skills training.

Organization of the Dissertation

This dissertation addresses the issues outlined above. The dissertation presents three papers that focus on social information processing (SIP) theory and its application in preventing aggressive behavior in childhood. The first paper reviews the SIP model and develops a general framework for applying SIP theory in behavioral interventions. The paper also discusses methodological issues in conducting SIP-based intervention. The second paper introduces social science and social work researchers to GPS methods for intervention research. This paper also reviews methodological limitations encountered in early dosage analyses and discusses challenges in conducting dosage analysis using GPS as well as areas for future research. An example is provided that applies a GPS method to a dosage analysis when dosage varies continuously. The third paper presents a dosage analysis of a SIP-based intervention, the *Making Choices* program. The evaluation of dosage effects uses a GPS method with continuous treatment. The paper discusses issues in implementing intervention programs in real-world settings.

REFERENCES: INTRODUCTION

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*, 1175-1184. doi:10.1037//0003-066X.44.9.1175

Conduct Problems Prevention Research Group (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648-657. doi:10.1037/0022-006X.67.5.648

Dodge, K. A. (1980). Social cognition and children’s aggressive behavior. *Child Development, 51*, 162–170. doi:10.2307/1129603

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology ,18*, 791-814. doi:10.1017/S0954579406060391

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology, 78*,107-126. doi:10.1037/h0031360

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, *5*, 163-183. doi:10.1007/s11292-009-9071-y

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology, 36*, 127-151. doi:10.1375/acri.36.2.127

Flannery, D. J., Vazsonyi, A. T., Liau, A. K., Guo, S., Powell, K. E., Atha, H., Vesterdal, W., & Embry, D. (2003). Initial behavior outcomes for the PeaceBuilders universal school-based violence prevention program. *Developmental Psychology*, *39*, 292-308. doi:10.1037/0012-1649.39.2.292

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

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. Y. (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*, 1045-1055. doi:10.1037/0022-006X.73.6.1045

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

Grossman, D., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K., Beland, K., …Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association*, *277*, 1605-1611. doi:10.1001/jama.277.20.1605

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp. 73-84). New York, NY: John Wiley. doi:10.1002/0470090456.ch7

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

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association, 99*, 854-866. doi:10.1198/016214504000001187

Imbens, G. W. (2000). The role of propensity score in estimating dose-response functions. *Biometrika*, *87*, 706-710. doi:10.1093/biomet/87.3.706

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology*, *150*, 327-333. doi:10.1093/oxfordjournals.aje.a010011

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. doi:10.1016/j.appdev.2003.08.002

Malti, T., Ribeaud, D., & Eisner, M. (2011). The effects of two universal preventive interventions to reduce children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40,* 677-692. doi:10.1080/15374416.2011.597084

Merrell, K., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly*, *23*, 26-42. doi:10.1037/1045-3830.23.1.26

Meyer, A. L., & Farrell, A. D. (1998). Social skills training to promote resilience in urban sixth-grade students: One product of an action research strategy to prevent youth violence in high-risk environments. *Education and Treatment of Children*, *21*, 461-488.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Developmental Psychopathology, 13,* 355-375. doi:10.1017/S0954579401002097

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N. P., Hancox, R., Harrington, H., ... Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*, 673-716. doi:10.1017/S0954579408000333

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. doi:10.1037/0022-006X.72.1.103

Rosenbaum, P. R. (1991). Discussing hidden bias in observational studies. *Annals of Internal Medicine, 115*(11), 901-905. Retrieved from http://annals.org/article.aspx?volume=115&page=901

Shure, M. B., & Spivack, G. (1988). Interpersonal cognitive problem solving. In R. H. Price, E. L. Cowen, R. P. Lorion, & J. Ramos-McKay (Eds.), *14 Ounces of Prevention: A casebook for practitioners* (pp. 69-82). Washington, DC: American Psychological Association.

Spivack, G., & Shure, M. B. (1974). *Social adjustment of young children: A cognitive approach to solving real-life problems*. San Francisco, CA: Jossey-Bass.

Wilson, S. J., & Lipsey, M. W. (2006). *The effects of school-based social information processing interventions on aggressive behavior. Part I: Universal programs*. Campbell Systematic Reviews. doi:10.4073/csr.2006.5

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building academic success on social and emotional learning: What does the research say?* New York, NY: Teachers College Press.

PAPER I

PROMOTING SOCIAL COMPETENCE AND PREVENTING AGGRESSIVE BEHAVIOR IN CHILDHOOD: A FRAMEWORK FOR APPLYING SOCIAL INFORMATION PROCESSING THEORY IN SCHOOL-BASED PREVENTION PROGRAMS

Over the past decades, the introduction and subsequent development of social information processing (SIP) theory has substantially advanced the understanding of the ways in which a child's cognitive operations can lead to aggressive behavior. Despite these advances in understanding the cognitive bases of aggression, applying SIP theory in intervention research to promote social competence and prevent aggressive behavior in childhood remains in a formative stage. Few programs have explicitly applied SIP theory in guiding curriculum design of a school-based intervention. Moreover, among the relatively few programs that have used SIP as a theoretical basis, the applications of SIP theory vary widely across those programs. Perhaps of even greater concern is the too frequent conflation of SIP with a traditional social problem-solving approach. To address these gaffs and gaps, this paper provides a general framework for applying SIP theory to school-based universal interventions. Key elements of the SIP model and subsequent debate are reviewed. A SIP-based intervention and a social problem-solving program are distinguished. Several key methodological issues in SIP research and intervention study are discussed.

Overview

Since the 1980s, social cognitive research has made a variety of significant advances in illuminating the impact of social cognition on behavioral responses in social interactions among children (e.g., Crick & Dodge, 1994; Dodge 1986; Fontaine & Dodge, 2006; Huesmann, 1988, 1998; Lemerise & Arsenio, 2000). These advances have centered on the ways in which distinct patterns of social information processing (SIP) or cognitive processes can lead to aggressive responses in social interactions among children. In developmental research, the SIP model has become a major theoretical model for understanding the cognitive mechanisms underlying aggressive behavior in childhood (Arsenio & Lemerise, 2010).

The SIP model describes specific cognitive processes that can be taught to children, and thus it has important implications for designing interventions to promote social competence and prevent aggressive behavior among children. However, the translation of the SIP perspective into practice—particularly school-based universal interventions—is in the formative stage with few applications to real world settings. Notably, only a few school-based universal programs have explicitly used the SIP model in guiding the design of their curriculum (e.g., Fraser et al., 2005; Meyer & Farrell, 1998). Moreover, not all researchers who have used SIP theory to design interventions appear to have recognized the unique contribution made by SIP theory. For example, when researchers have applied multiple theories to curriculum design, including social learning theory (Patterson, 1986), cognitive-excitation approaches (Zilman, 1979), cognitive scripts (Huessman, 1988), and SIP theory (Crick & Dodge, 1994; Dodge, 1986), some have considered the implications of the SIP model as similar to those of the other theories (Meyer & Farrell, 1998). Nevertheless, perhaps an issue of greater concern is that researchers frequently confuse the traditional social problem-solving (SPS) approach (D'Zurilla & Goldfried, 1971; Shure & Spivack, 1988; Spivack & Shure, 1974) with the SIP framework (e.g., Frey, Hirschstein, & Guzzo, 2000; Wilson & Lipsey, 2006). Essentially, these approaches should be differentiated in that the SPS approach does not treat behavior as a function of a sequenced cognitive process, such as the process specified by the SIP model. However, when lessons were constructed within a traditional SPS framework and used a 5-step problem-solving strategy (i.e., identify the problem; brainstorm solutions; select, plan, and try the solution; evaluate if the solution worked; and decide what to do next), researchers regarded the 5-step problem-solving strategy as addressing each of the SIP processes (e.g., Frey, Hirschstein, & Guzzo, 2000).

This conflation of SPS with SIP also has led the authors of systematic reviews of school-based universal SIP interventions to identify many programs as SIP interventions although these programs scarcely mention SIP theory (e.g., Bosworth, Espelage, DuBay, Daytner, & Karageorge, 2000; Denham & Burton, 1996; Forness et al., 2000; Lynch, Geller, & Schmidt, 2004; Nelson & Carson, 1988; Shapiro, Burgoon, Welker, & Clouch, 2002). Wilson and Lipsey (2006), for example, conducted a review to examine the effects of social skills training programs on aggressive and disruptive behavior among school-aged children. For their review, they broadly defined a SIP intervention as any program that provided training on one or more of the SIP steps. However, most of the programs they identified under this broad definition as SIP interventions actually used a traditional SPS approach (e.g., Frey, Hirschstein, & Guzzo, 2000; Sawyer et al., 1997; Work & Olsen, 1990). Further, although the cognitive skills emphasized by an SPS approach are not specific to SIP, they were considered training on the SPS skills equivalent to addressing cognitive skills deficits corresponding to SIP steps.

In sum, despite advances in understanding the contribution of cognitive factors to aggressive behavior in childhood, applying the SIP perspective to interventions remains in a formative stage. The unique contribution of SIP has not been fully recognized by intervention researchers in designing curriculum. Moreover, conflating traditional SPS approaches with SIP interventions is common.

This paper provides a framework for applying SIP theory as a guide for program or, as it is called in school-based research, curriculum design. The framework was formulated by incorporating the essential ideas of SIP and the subsequent research and dialogue regarding the SIP model. The description of the framework is followed by discussion of the distinct characteristics of the SPS and SIP models. Finally, methodological challenges in SIP research and intervention study are discussed.

Two General Models of SIP

During the 1980s, two general models of information processing were introduced: one by Dodge (1986) and one by Huesmann (1988). Both were subsequently reformulated to explain how humans acquire and maintain aggressive behavior (Crick & Dodge, 1994; Huesmann, 1998). Both models elaborated a sequential process of cognitive tasks that individuals undertake in a social situation. However, these models differed in critical ways. In the revised SIP model formulated by Crick and Dodge (1994), social cognitive processing is thought of as an *on-line* (i.e., real-time) and conceptual process; whereas in the Hussmann model, social cognitive processing is thought of as schema based (i.e., script based) and an automatic process. Crick and Dodge focused on the immediate effects of cognition on behavior in a specific instance. In contrast, Huesmann focused on scripts and the acquisition and retrieval of those scripts.

The scholarly community responded to the two SIP models in contrasting ways. The Crick and Dodge model attracted wide attention in the social development sphere and stimulated considerable scholarly thought and investigation, whereas the Huesmann model was given substantially less attention. The lack of either positive or negative feedback on his model led Huesmann to comment that his theory had been "missed by many developmental researchers on social adjustment" (Huesmann, 1998, p. 89).

The tepid reception given to Huesmann's model has a possible historical explanation. Traditionally, social cognitive research has focused on off-line (or latent) cognitive structures such as values (e.g., Boldizar, Perry, & Perry, 1989; Nucci & Herman, 1982; Turiel, 1983), schemata or scripts (Abelson, 1981; Huesmann, 1988; Schank, 1977), and beliefs (e.g., Huesmann & Guerra, 1997), and the ways in which cognitive structures are acquired and affect interpersonal behavior of children. Although research on latent mental structures has made notable contributions to the understanding of social cognition and behavior of children, this traditional research approach has been unable to explain how cognitions affect immediate behavioral responses in particular situations (Fontaine, 2008). Focusing on latent structures (i.e., schemas, scripts, and beliefs), Huesmann's model (1988, 1998) was more in line with the research tradition of off-line cognitive structures. Unlike Huesmann's model, Dodge's (1986) initial model (as well as Crick and Dodge's later, reformulated model) addressed a gap in traditional social cognitive research by providing a framework for understanding the immediate effects of cognition on the behavior of children. Because the

Crick and Dodge model addressed the existing gap in cognitive understanding, rather than taking the traditional approach, their model garnered greater attention.

The considerable scholarly thought and investigation inspired by the Crick and Dodge SIP model accelerated a trend toward focusing on on-line processing in the field of developmental psychopathology. Within this field, the Crick and Dodge SIP theory has been accepted as a major theoretical framework for understanding the ways in which cognitive factors can lead to aggression in specific situations (Lansford et al., 2006). Moreover, in Crick and Dodge's reformulation of the SIP model, more attention was paid to the interaction of on-line processing and latent structures. Indeed, the unique contribution of Huesmann's model (1988, 1998) may be better understood in the general framework offered by Crick and Dodge. The reformulated SIP model provides insights into how latent structures affect on-line processing and shape behavior of children. Given these reasons, this study focuses on Dodge's (1986) initial SIP model and the reformulated Crick and Dodge (1994) SIP model.

Although Dodge's (1986) initial SIP model garnered much attention and had substantial influence on social cognition research, the model was controversial (e.g., Arsenio & Lemerise, 2004; Gottman, 1986; Sutton, Smith, & Swettenham, 1999). Because the model has been refined and modified since its introduction, an overview of the history of the model is provided, including a discussion of the contentious issues that are relevant to the design of SIP-related interventions. A general framework for applying SIP is then developed by incorporating the initial SIP model and subsequent debate about the model.

Dodge's SIP Theory of Aggression

Dodge's (1986) SIP theory proposed that when children are in a social situation they are faced with an array of cues from which they have to choose and then process by engaging in specific cognitive steps before enacting a behavioral response. Crick and Dodge (1994) reformulated this initial SIP model and the revised model has become the dominant SIP model. The Crick and Dodge SIP model proposed that behavioral responses to social situations were the end product of cognitive processing that occurred in five overlapping steps. These cognitive steps include Step 1, encoding of external and internal cues; Step 2, interpreting or cognitive representation of those cues; Step 3, choosing and clarifying a goal; Step 4, selecting or constructing a response; and Step 5, performing the response decision (for reviews, see Crick & Dodge, 1994, 1996; Dodge, 2006). The revised SIP model also posited that these on-line (i.e., real time) processing steps, including the final step of enacting the selected behavioral response, were influenced or guided by latent mental structures (e.g., social schema, scripts, and social knowledge) that the child developed from accumulated memories of events and experiences. Similarly, engagement in each step was conceptualized as having the potential to bring about changes or revisions to the latent cognitive structures.

The SIP model is formulated as a global framework that represents cognitive operations underlying child behavior. However, the model's primary application has been to understanding aggressive behavior in children, which is defined as “behavior that is aimed at harming or injuring another person or persons” (Parke & Slaby, 1983, p. 550). Many empirical studies have demonstrated the relationship of patterns in cognitive processing at each SIP step with aggressive behaviors. Specifically, research has shown that as compared with their nonaggressive peers, aggressive children encode fewer and less-benign social cues (Step 1; Dodge & Newman, 1981; Gouze, 1987; Strassberg & Dodge, 1987); attribute more hostile intentions to others (Step 2; Feldman & Dodge, 1987; Dodge, Bates, & Pettit, 1990); select goals that are more likely to damage relationships (Step 3); generate fewer response options and develop responses that are less prosocial (Step 4; Pettit, Dodge, & Brown, 1988); and evaluate aggressive responses more favorably and expect more positive outcomes from aggressive behavior (Step 5; Dodge et al., 1990; Perry, Perry, & Rasmussen, 1986).

Crick and Dodge's (1994) SIP model emerged from a research tradition that examined social cognition of children based on the premise that social cognitions were the mechanisms leading to social behaviors. Earlier work using this approach focused on global cognitive constructs such as perspective taking, role taking, and referential communication (e.g., Flavell, Botkin, Fry, Wright, & Jarvis, 1968; Selman, 1971). Early tests using global cognitive constructs to predict social behavior produced mixed findings (e.g., Shantz, 1975, 1983). The 1970s ushered in the introduction of theories of information processing by researchers such as Newell and Simon (1972). Rather than a global cognitive construct approach, the new theories focused on specific components or steps of on-line cognition. This perspective of "real-time cognition" quickly gained popularity, and led to major changes in empirical and theoretical approaches to the study of social cognition in children. Crick and Dodge were among the major contributors in these new approaches. By specifying the information processing steps in which children engage, the Crick and Dodge (1994) SIP model constituted a substantial advancement in the understanding of social adjustment of children. Because it described specific processes that can be taught to children, their model has served as an important guide in designing interventions for use with social maladjustment in children (e.g., Conduct Problems Prevention Research Group, 1992).

Debate Regarding the SIP Model: Implications for Intervention

Despite its wide appeal, the Crick and Dodge (1994) model was criticized for ignoring the role of emotion, being value blind (Arsenio & Lemerise, 2004; Sutton, Smith, &

Swettenham, 1999), and assuming a homogeneity of SIP deficits among aggressive children (Sutton et al., 1999). Clarifying these issues has been important in further specifying the SIP model and in providing guidance for the design of interventions.

Role of Emotion

Criticism of Crick and Dodge's (1994) SIP model for neglecting the influence of emotion on the cognitive processing of social information had its roots in work by Gottman (1986). Dodge (1991) responded to his critics by proposing that emotions are integral to each SIP step "in that emotion is the energy level that drives, organizes, amplifies, and attenuates cognitive activity and in turn is the experience and expression of this activity" (p.159). In reformulating the SIP model, Crick and Dodge (1994) acknowledged that emotion was relatively neglected in the initial model, and provided modest explanations of how emotion and cognition interact at each SIP step. Given the added attention to emotion in the reformulated SIP model, Dodge and Rabiner (2004) rejected Arsenio and Lemerise's (2004) criticism that emotion was ignored in the SIP model. They argued that "processing is meant to be entirely emotional" (Dodge & Rabiner, 2004, p. 1006). However, compared with Lemerise and Arsenio's (2000) more comprehensive treatment, Crick and Dodge's articulation was modest and more limited in explaining the role of emotion in cognitive processing of social information in childhood.

In contrast, Lemerise and Arsenio (2000) provided a fuller explication of the role of emotion in SIP. For example, whereas Crick and Dodge (1994) suggested that a child enters a social situation with a combination of "biologically limited capabilities and a database of memories" (p. 76), Lemerise and Arsenio (2000) expanded on that description by proposing that a critical component of biological predisposition was the child's emotional style or emotionality (Eisenberg & Fabes, 1992; Rothbart & Derryberry, 1981). In addition, a child's representations of his or her experiences also include affective components. Specifically, children were argued to vary in the intensity with which they experience and express emotions as well as in their skills for regulating their emotions. The intensity of emotions and a child's regulatory capacities were conceptualized as influencing each SIP step.

Although relatively little research has been conducted on SIP and emotion, previous research has both directly and indirectly provided support for Lemerise and Arsenio's perspective. Specifically, the intensity of emotions and the capacity to regulate emotions influence which of the many cues are noticed in a social situation and what meaning is attributed to the situation (Steps 1 and 2; Casey, 1996; Casey & Schlosser, 1994). Children who are overwhelmed by their own or other's emotions may choose avoidant or hostile goals (Step 3; Eisenberg & Fabes, 1992; Eisenberg, Fabes, Nyman, Bernzweig, & Pinuelas, 1994; Saarni, 1999; Sroufe, Schork, Motti, Lawroski, & LaFreniere, 1984). Furthermore, children who experience strong emotions but lack the skills to regulate their emotions in challenging situations may be overwhelmed and become too self-focused to generate a variety of responses and evaluate those responses from the perspectives of all parties involved in the situation (Steps 4 and 5; Eisenberg et al., 1994; Saarni, 1999). Finally, children who are in conditions of high emotional arousal are likely to resort to using inflexible approaches to situations (Step 6; Casey, 1996; Casey & Schlosser, 1994; Saarni, 1999). In sum, high emotionality and poor emotional regulation are likely to produce deficits in SIP that contribute to aggressive behavior (e.g., Murphy & Eisenberg, 1997; Pakaslahti, 2000).

Lemerise and Arsenio (2000) advanced the field. However, they did not discuss how the influence of emotion on SIP might vary in relation to important variables such as social context (e.g., peer group entry and provocation), gender, age, and type of aggressive behavior. Nevertheless, emotion is now widely acknowledged as an integral part of SIP, and therefore, emotional regulation training has emerged as an element of SIP-based interventions. In particular, strategies that are aimed at both increasing the awareness of emotions and enhancing capacity for emotion regulation in children are considered elemental in the design of SIP interventions.

Influence of Values on Latent Cognitive Structures

Another criticism of the Crick and Dodge (1994) model referred to the value-free nature of the SIP model (Arsenio & Lemerise, 2004; Sutton et al., 1999). Among these critics, Arsenio and Lemerise (2004) created a theoretical model that integrates SIP with a moral domain model. This elaboration provides greater specificity regarding the ways in which latent mental structures (e.g., moral knowledge structure) interplay with on-line SIP.

Moral knowledge structure is likely but one of several latent knowledge structures with the potential to influence processing. Indeed, Crick and Dodge (1994) proposed several potentially influential latent knowledge structures, including schemata (Mandler, 1979; Winfrey & Goldfried, 1986); scripts (Schank & Abelson, 1977); internal working models (Bowlby, 1969, 1973, 1980); and cognitive heuristics (Einhorn & Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982).

An internal *working model* is a concept drawn from attachment theory. It was originally defined as mental representations of the self, attachment figures, and the relationship between the two (Bowlby, 1980). Later theorists have proposed that internal working models are organized in a hierarchical fashion, with the lowest level of the hierarchy composed of specific scripts (e.g., “My mother comforts me when I get hurt”) that are generalized from repeated experiences with attachment figures. The higher levels of the hierarchy are derived from lower levels and are composed of increasingly general schemas regarding attachment figures and the self (e.g., "My mother cares for me when I need her"; Bowlby, 1980; Brethertion, 1985). In the context of understanding aggressive behaviors in childhood, an internal working model implies that in addition to a model of self, children have internal working models of their peers that are generalized from experiences with peers and assumptions about peers that have been abstracted from events.

In addition to internal working models, cognitive heuristics have been identified as having the potential to influence SIP of children. Heuristics are simple, efficient rules that people use to make decisions, come to judgments, and solve problems. These experience-based precepts are typically used when a person is facing a complex problem or making a choice with incomplete information (for reviews, see Einhorn & Hogarth, 1981; Kahneman et al., 1982). For example, a child is punched in the back and turns around to face two possible aggressors: one child is tall and strong, and the other is small and frail. The child who was attacked might use “bullies have strong arms” as a heuristic, or rule-of-thumb, and assume the tall, strong child must be the person responsible for the attack.

Moreover, Dodge and Rabiner (2004) argued that a moral knowledge structure might provide less explanatory power than other latent knowledge structures, such as working models, in understanding the processes leading to aggressive behavior in childhood. For example, when children experience harm, particularly when the intent cues are ambiguous, their internal working models of peers are more likely to be activated than moral knowledge when making intent attribution. Children are less likely to make an attribution of hostile intent based on their judgment of whether a peer’s behavior represents a moral transgression.

At the goal clarification and selection stage, the relative emphasis of children on relational versus instrumental goals is also influenced more powerfully by their working model of peers than by their moral knowledge structure. In general, when children find comfort, pleasure, and satisfaction through peer relationships, specifically with the *peer provocateur* (i.e., the person who incites or stimulates a child to action), they are likely to view relational goals more favorably. At the response-generation and response-selection steps, Arsenio and Lemerise (2004) suggested that the underlying moral knowledge structures exert strong selective pressure for certain choices. Dodge and Rabiner (2004) argued that at this stage they "expect the integration with moral domain theory to be most fruitful" (p. 1006). However, Dodge and Rabiner also proposed that "decisions to engage in certain behaviors will depend on how a child expects a particular response will affect future relations with a peer as well, and this judgment will be influenced by the child's working model of relationships" (p. 1006).

Dodge and his colleagues' elaboration regarding the strong explanatory power of the working model of peers has been largely accepted. However, one important factor was missing in their discussion: the perception of harm in a provocation. In a situation of provocation, the perception of potential harm can be translated to questions such as to what extent the provocation matters and how difficult it would be to recover from the harm. It is easy to imagine how children who generally favor relational goals and nonaggressive behavior could be provoked to respond aggressively when harmed. Therefore, it seems likely that the perceived level of harm associated with a provocation interacts with working models of peers in processing social cues. The perceived level of harm is important in that it provides motivation as well as justification for aggressive behavior.

Despite the limitations noted above, the debate on the relative importance of latent mental structures and how children translate structural knowledge into behavioral responses has contributed to understanding aggressive behavior in childhood. Additionally, it has provided important guides for the design of interventions. Particularly, given the potential influence of internal working model of peers, it is important to develop strategies to change perceptions of peer relationships among children. Activities that can provide children with positive peer experiences are expected to be beneficial in altering processing biases that lead to aggressive behavior.

In contrast, the way in which latent structural knowledge is activated at different processing steps remains largely unexplained, and is an area that needs continued research. The existence of multiple knowledge structures and their contextualized application require work to develop a clearer specification of the mechanism underlying aggressive behavior in childhood.

Reactive and Proactive Aggression and SIP

Aggressive behavior is multidimensional. In the literature, the construct *aggression* is multi-defined as being direct, indirect, overt, relational, social, physical, verbal, nonverbal and nonphysical, reactive, and proactive (Camodeca & Goossens, 2005; Crick, Casas, & Nelson, 2002; Crick & Dodge, 1996; Crick & Grotpeter, 1995; Crick, Grotpeter, & Bigbee, 2002; Dodge & Coie, 1987; Grotpeter & Crick, 1996; Rigby, 1996). Bullying is also discussed as a form of aggression, and most often is identified as a form of proactive aggression (Sutton et al., 1999; Baldry & Farrington, 2007). Applying these various distinctions to aggressive behavior is not only confusing but also the subject of ongoing debate.

A thorough treatment on all subtypes of aggression is beyond the scope of this paper. This paper focuses on reactive and proactive aggression as related to SIP patterns. Compared to other subtypes of aggression, reactive and proactive aggression have received relatively more attention in SIP studies. *Reactive aggression*, which has theoretical roots in the frustration-aggression hypothesis (Berkowitz, 1963; Dollard, Doob, Miller, Mowrer, & Sears, 1939), is described as an angry, defensive, retaliatory response to provocation. In contrast, *proactive aggression* is characterized as unprovoked, deliberate, goal-directed behavior used for coercion (Dodge & Coie, 1987; Hubbard et al., 2002). The theoretical roots of proactive aggression are in social learning theory (Bandura, 1973), which postulates that aggression is an acquired behavior controlled by reinforcements.

The SIP model has been criticized for treating aggressive children as a homogeneous group in terms of social skills deficits (Sutton et al., 1999). Sutton and colleagues suggested that, unlike reactive aggressors, proactive aggressors perceive and interpret social cues accurately, but differ from nonaggressive children in their patterns of goal selection, response strategy generation, and response decisions. Although Sutton and colleagues (1999) used the SIP framework to differentiate reactive and proactive aggression, they claimed, "the model as a whole may apply more to reactions and reactive aggressors than actions or proactive aggressors" (p. 122).

These criticisms are arguable. First, Crick and Dodge (1994) never used the term *social skills deficit* to conceptualize SIP patterns that lead to aggressive behavior. Evaluating aggression and its consequences positively cannot be simply explained by skills deficits. In response to Sutton and colleagues' criticism, Crick and Dodge (1999) argued that "The SIP framework … does not require that aggressive behavior occurs as a function of processing deficits…Rather the key formulation of the SIP framework is that chronic processing styles account for chronic patterns in aggression" (p. 128).

Second, early research in child aggression was characterized by its relative neglect of the distinction between reactive and proactive aggression among children (Dodge & Coie, 1987; Hartup, 1974; Rule, 1974). Dodge and his colleagues were among the early contributors who attempted to distinguish between reactive and proactive aggression among children. Using the SIP framework, Dodge and Coie (1987) found that reactive aggressors had hostile biases and deficits in perceiving and interpreting social cues. Later, Crick and Dodge (1996) found that proactive aggressors tended to select instrumental social goals rather than relational goals, and evaluated aggression and its consequences in relatively positive ways. These findings have demonstrated that the SIP framework applies to both reactive and proactive aggression, and the processing patterns at each step distinguished between reactive and proactive aggression.

Given this evidence, some concerns of critics seem ill founded. Intervention researchers should move forward to focus on developing strategies to alter different processing patterns that lead to different types of aggression. Specifically, skills training to enhance perceiving and interpreting social cues might underpin interventions focused on reactive aggression, whereas training to alter patterns of positive evaluation of aggression and its consequences might underpin interventions focused on proactive aggression.

Framework for Designing SIP Interventions

The SIP model has contributed substantially to the current understanding of the mechanisms underlying aggressive behavior of children. The debates on the SIP model have expanded knowledge regarding the role of emotion, values, and other latent mental structures in social interactions. Based on the SIP model and incorporating the subsequent discussions of the model, I have developed a framework for designing SIP interventions. This framework proposes that the design of a typical SIP-based intervention should aim to alter the biases in processing patterns at each SIP step that lead to aggressive behavior in childhood. Strategies to enhance emotional competence and to refine latent cognitive structures of children related to SIP are considered integral components of a SIP intervention.

In SIP theory, aggressive behavior in childhood is understood as a function of patterns of biased processing that can occur at each of the SIP steps. Therefore, a logical implication of this model is to provide strategies to alter processing patterns of children that can lead to aggressive behavior at any of the SIP steps. As such, it is critical to apply the SIP theory in a comprehensive way when designing SIP-based interventions rather than applying separate elements of the SIP theory. Interventions that do not target comprehensive processing patterns that encompass all the SIP steps should be avoided.

Of course, many theories have contributed to the understanding of aggressive behavior of children and, therefore, provide valuable implications for designing interventions. The variety of theories attempting to explain behavior of children and the variation in applying SIP theory lead to important questions: Should the SIP model be an additional, supplementary theoretical source from which researchers draw a few implications similar to those from other theories? Or, should the SIP model be the primary theoretical base for designing curriculum and serve as the organizing framework within which other relevant theories can be applied in an integrated manner?

The answer resides in the unique contribution and utility of the SIP model. The SIP model specifies the cognitive steps that children engage in before enacting behavior when faced with social situational cues. As compared with global constructs (e.g., perspective taking, role taking, referential communication), the specific processing components (i.e., steps of on-line cognition) have been shown to be more predictive of social adjustment of children (Crick & Dodge, 1994). Moreover, by accounting for both on-line and latent mental structures, the SIP model has provided a comprehensive framework for studying social behavior of children. The introduction and development of the SIP model has been widely acknowledged for providing substantial advances in the understanding of social adjustment in childhood. Indeed, the SIP model has had a major impact on the overall direction of research on the social development of children (Huesmann, 1998). Given the evidence of the utility of the SIP model, this model should be a primary theoretical base for designing curriculum for promoting social competence and preventing aggressive behavior among children.

Moreover, by linking biologically limited capabilities of children with their on-line processing and latent mental structures, the SIP framework can serve as an organizational framework for integrating a variety of theories relevant to the development and maintenance of social behaviors of children. For instance, social learning theory has been used to guide the development of many programs. Within a SIP framework, social learning theory can be applied to explaining how certain latent mental structures are formed. Although a detailed discussion on this issue is beyond the scope of this study, the existing evidence suggests that integrating theories relevant to social adjustment within a SIP framework would substantially advance the knowledge base regarding cognitive patterns, and especially patterns of socially maladjusted children. In addition, an integrated SIP framework would advance the understanding of the risk and protective factors that contribute to social maladjustment and provide new insight toward the development of treatment strategies for maladjusted youth.

Distinguishing SIP-Based Interventions and SPS Interventions

The conflation of an SPS approach with a SIP framework can pose a major obstacle in translating the SIP theory into an intervention study. Therefore, distinguishing a SIP model from an SPS approach represents a crucial step in applying SIP theory to the design of an intervention to promote social competence and prevent aggressive behavior in children.

A traditional SPS approach emphasizes three cognitive skills: (a) *alternative thinking*, that is, the ability to generate multiple alternative solutions to interpersonal problems; (b) *consequential thinking*, or the ability to anticipate immediate and long-term consequences of actions; and (c) *means-to-ends thinking*, the ability to create a plan of specific actions to attain a goal and to recognize and deal with obstacles (D'Zurilla & Goldfried, 1971; Shure & Spivack, 1988; Spivack & Shure, 1974). Typical SPS interventions provide training in problem-solving skills using a five-step approach. The five steps are (a) identifying the problem; (b) brainstorming solutions; (c) selecting, planning, and trying the solution; (d) evaluating if the solution worked, and (e) deciding what to do next (e.g., Frey, Hirschstein, & Guzzo, 2000).

Shown in Table 1, an SPS approach differs fundamentally from a SIP framework in that SPS does not treat behavioral responses as a function of sequential cognitive operations illustrated by the SIP model. Consequentially, the SPS five-step problem-solving strategy does not fully address processing patterns at each SIP step that can lead to aggressive behavior. For instance, some aggressive children have been characterized as selectively attending to social situational cues, encoding fewer cues, making hostile attribution, and favoring instrumental goals over relational goals. These issues were not directly addressed by the five-step problem-solving strategy, although SPS skills training might be beneficial in changing processing bias of children at some SIP steps.

Table 1.1
Comparison of a Typical SIP Program and a Typical SPS Program

SIP: 6-Step Cognitive Process	SPS: 5-Step Problem Solving Strategy
Identify the cues	Identify the problem
Interpret the cues	Brainstorm solutions
Set up goals	Select, plan, and try the solution
Access or construct responses	Evaluate if the solution worked
Evaluate and select a response	Decide what to do next
Enact a response	

Methodological Issues in SIP Research and Intervention Study

Numerous methodological factors contribute to the validity of findings, including research design, sample selection, measurement precision, data collection, and selection of appropriate analytic methods. The following discussion on methodological challenges in SIP research focuses on issues related to data collection. Data collection comprises two major aspects: determining what variables should be measured, and determining what means should be used to measure those variables.

What to Measure? Variables That Should Not Be Ignored

One important task in intervention research is to answer an expanded version of Gordon Paul's (1967, p. 111) "ultimate" question: How much of which intervention by whom is most effective for which participants with what type of problem? Given the impact of program participation on all participants and the substantial costs of intervention programs, it is impossible to overstate the importance of providing unbiased and detailed information on program effectiveness.

A program usually comprises multiple components, with each component consisting of a variety of active elements, and each contributing to some extent to program outcomes. In addition, delivery, training, and organizational variables all contribute to outcomes (Elias, 1994). It is crucial to identify and assess these content, process, fidelity, and dosage variables that potentially influence outcomes.

In general, evaluation studies have taken a comprehensive approach by examining variables from a range of perspectives, including theoretical, training, implementation, and environmental. However, previous studies have failed to capture two important factors: the effects of teacher characteristics (e.g., teaching style) and a student's level of classroom involvement (e.g., interest, investment). These two factors are related in that, to a certain extent, the teaching effectiveness accounts for students' classroom involvement.

Importance of teacher characteristics. Studies have found that teachers' personal characteristics and their instructional strategies were among the key factors associated with successful education (e.g., Pressley et al., 1992). This assertion is particularly salient in social skills education. In addition to the teacher's influence on student classroom involvement in regular academic classes, those teachers who conduct social skills training also serve as important role models for students' social learning. A teacher who uses harsh, judgmental, or derogatory language in interactions with students is unlikely to be successful in teaching students to develop prosocial behavior. The critical role of the teacher in social-emotional skills training has been widely acknowledged by intervention researchers. For example, Frey and colleagues (2005) explicitly viewed the teacher as a "second source of learning" (p. 195) in addition to the curriculum. Meyer and Farrell (1998) emphasized the importance of the teacher as a role model and stated, “The purpose of this curriculum is for a valued adult role model to teach students knowledge, attitudes, and skills that promote non-violence and resilience” (p. 11). However, this level of acknowledgement has not been successfully translated into an emphasis on investigating variables such as teachers’ characteristics or teaching quality in social intervention studies.

Student classroom involvement. Another variable missing from evaluations of social-emotional skills training programs is students’ classroom involvement. Researchers have made advances in developing strategies to optimize class participation by all students. Unfortunately, the actual involvement of students has not been assessed. Finding the optimal dosage of program curricula has been widely stressed by intervention researchers. However, little attention has been paid to differentiating the dosage delivered or program exposure and the dosage taken up by students. The quality of teaching and the actual involvement of students are two of the important variables that account for the dosage that students take up. Thus, research efforts to investigate the effects of teacher characteristics and student involvement are warranted.

How to Measure? Issues Regarding the Validity of SIP Measures

Hypothetical scenarios. Assessing cognitive operations associated with aggressive behavior is a challenging endeavor. The latent nature of cognitive processing constrains the use of direct observation. An alternative to observation is a laboratory approach, but this approach is cost prohibitive and often lacks feasibility, particularly in universal intervention settings. Hence, using hypothetical situations to identify SIP processing patterns often remains the best choice for measuring SIP variables (e.g., Arsenio, Adams, & Gold, 2009;

Crick & Werner, 1998; Fraser et al., 2005). However, using hypothetical scenarios to elicit cognitive processing and behavioral responses has an inherent limitation: emotions provoked by hypothetical scenarios might differ substantially from the emotions experienced in real-life situations. Moreover, the need for comprehensive considerations of multiple factors poses considerable challenges in designing instruments to measure SIP variables. These factors include the types of situation (e.g., instrumental vs. relational), nature of intent (e.g., hostile, benign, accidental, and ambiguous), types of aggression (e.g., reactive vs. proactive), and the characteristics of the provocateur (e.g., aggressive vs. prosocial). In addition to the use of hypothetical situations, SIP measures used in previous studies were generally limited by the exclusive use of ambiguous intent as stimuli and by a lack of information about the provocateur (e.g., the provocateur's general characteristics and relationship with the victim).

Personal involvement. The limitation imposed by the use of hypothetical scenarios is related to the issue of personal involvement. Most studies ask children to imagine themselves to be the victim in a hypothetical scenario and to respond to a specific provocation. This approach is substantially limited by the extent to which a child might experience the emotion of the "real" victim. Indeed, one recent study with preschool children found that most children refused to take the role of the child in the videos if the video portrayal used a child of the opposite sex (Schultz et al., 2010).

Emotion is an integral part of SIP. Emotion has been recognized as "the energy level that drives, organizes, amplifies, and attenuates cognitive activity and in turn is the experience and expression of this activity" (Dodge, 1991, p. 159). Discussed previously, the intensity of emotions can influence the way in which social information is processed at each SIP step (e.g., Casey, 1996; Casey & Schlosser, 1994; Eisenberg & Fabes, 1992; Lemerise &

Arsenio, 2000; Saarni, 1999). However, in hypothetical situations, it is unlikely that children will experience the same intensity of emotion as they do when facing the same situation in real life. Consequently, the cognitive process and behavioral responses elicited by the hypothetical stimuli might not be the same as responses in real-life situations, and therefore, could be biased. In addition, when providing answers to questions about their use of aggressive behavior, children are particularly vulnerable to social desirability, and consequently, might not give answers that truly represent their usual behavior.

One strategy to increase the level of personal involvement and minimize the effects of social desirability might be to ask a child to respond on behalf of the victim (e.g., “What do you think he or she will say or do?”). This strategy was used by Schultz and colleagues (2010) when they found children refused to pretend to be the child of the opposite sex in the videos. The shield of representing the thoughts and actions of the victim may actually more faithfully mirror the participant’s patterns of cognition and behavior.

Ambiguous stimuli and accidental stimuli. SIP measures used in previous studies are also characterized by the exclusive use of ambiguous situations to elicit responses in intention-cue detection and intent attribution (e.g., Arsenio et al., 2009; Crick & Dodge, 1996). Differences in intention-cue detection accuracy and tendency to make hostile attribution are found between subtypes of aggression (i.e., reactive and proactive) under both ambiguous stimuli and accidental stimuli (Dodge & Coie, 1987). This finding implies that in measuring encoding and attribution patterns, researchers should present both ambiguous and accidental stimuli. The absence of accidental stimuli could affect the validity of the measurement.

Lack of provocateur information. Another general limitation is related to the missing information regarding the provocateur, including the provocateur's general characteristics and relationship with the victim. Dodge and Rabiner (2004) suggested that when a child experiences harm—particularly when the intent cues are ambiguous—his or her internal working model of what peers are generally like is activated in making intent attribution and subsequent processing. Expanding on this notion, it is proposed that, when the victim is familiar with the provocateur, the provocateur's characteristics and his or her relationship with the victim influence – at least in part –the ways in which the victim makes intent attribution and response decisions. Indeed, it is easy to understand that a child is likely to respond differentially to a provocation made by a caring friend than to a provocation by a mean peer. One early SIP study (Milich & Dodge, 1984) used peer-nominated aggressive boys as the antagonists in hypothetical stories that were read to each participant, adding the the provocateur's information. However, a similar strategy has not been used in subsequent studies, which, on balance, might be related to concerns about human participant protection and research ethics.

Though less general, many studies have been characterized by a lack of information about the provocateur's facial and voice expressions (e.g., Arsenio et al., 2009; Crick & Dodge, 1996; Crick, Grotpeter, & Bigbee, 2002; Crick & Werner, 1998; Dodge et al., 1999). Presumably, facial and voice expressions are important elements in processing cues. Indeed, social-skills training generally encompasses identification of facial and voice cues. To present this information in uniform, valid ways, facial expressions can be presented with videos, photographs, and drawings whereas vocal expressions can be reproduced with audio recordings or via enactment by a research assistant. Presumably, video would be an effective way of providing detailed, realistic, and replicable facial and vocal cues. However, a meta-analytic review found the strongest association between hostile attribution of intent and aggressive behavior was in studies that used either audio presentations or text presentations of stories (Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). One explanation for this unexpected finding is related to personal involvement. Video presentations might make it difficult for participants to imagine themselves as the person "you don't know in a blue T-shirt on a TV" (Orobio de Castro et al., 2002, p. 929). Discussed previously, this issue related to videos might be partially addressed by asking participants what the child on the video might say or do.

An alternative way to present both facial and vocal cues is to use pictures with audio. Orobio de Castro and colleagues' findings (2002) suggested that picture presentation alone was associated with the smallest effect size, and that text presentation had a similar effect size as audio presentation. However, such findings do not preclude the use of combined audio-visual presentations. First, in their meta-analysis, only two studies used either pictures-only or text-only presentations. Second, no comparison has been done using the three options of pictures with audio, text and audio, and text only. Given the importance of facial and vocal cues in processing social information, studies might be limited by not providing all such information.

Both basic SIP studies and SIP intervention research are constrained by the quality and difficulty of measurement. Developing instruments to measure SIP variables is challenging given not only the nature of cognitive processing but also the variety of elements in different aspects of processing.

Conclusions

Social competence developed in childhood is a critical ability of an individual. It is related to a variety of developmental and adult outcomes. Failure to develop social competence is associated with negative developmental outcomes such as peer rejection and aggressive behavior (e.g., Smith, 2001; Trentacosta & Izard, 2007), and negative long-term socioeconomic outcomes (e.g., Heckman, 2008). Therefore, promoting social competence and preventing aggressive behavior in childhood are crucial areas for interventions. By specifying cognitive operations underlying behaviors of children, SIP theory has important applications for designing interventions to improve social competence and prevent aggressive behavior in childhood.

The SIP model has been widely acknowledged for providing a comprehensive understanding of the social adjustment of children. Both theoretical reasoning and empirical evidence have demonstrated the link between processing patterns at each SIP step with behavior acts. The relation of SIP patterns with behavior should lead to a comprehensive approach in applying SIP models to intervention. Although the extent to which the potential of the SIP model can be realized depends on numerous factors, a fundamental element of any effort should be designing an engaging program that provides social and emotional skills training aimed at altering SIP patterns associated with social maladjustment. Using SIP theory to modify training strategies within a traditional SPS framework or focusing only on selected SIP steps should be avoided.

Moreover, by linking latent cognitive structures with on-line processing, the SIP model provides the potential for integrating a variety of behavioral theories within the SIP framework. Given the comprehensive nature of the SIP model, there should be a change in the training paradigm from a traditional SPS model to a SIP-based intervention. Three decades ago, Ladd and Mize (1982) called for a precise and unified model of social skills training. Their voice is echoed here in a call for a comprehensive application of the SIP perspective.

Notwithstanding, SIP-based research faces many challenges. One challenge is the complexity of developing instruments to collect valid and reliable information on cognitive skills. Previous studies have identified several limitations regarding instruments and data collection, including the use of hypothetical situations, omitted accidental stimuli, and the lack of information on the provocateur. Because of the fundamental nature of measurement, additional research to address these issues is critically important. While challenging, addressing these issues is one of the emerging opportunities associated with the advances in cognitive research.

REFERENCES: PAPER I

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36*, 715-729. doi:10.1037/0003-066X.36.7.715

Arsenio, W. F., Adams, E., & Gold, J. (2009). Social information processing, moral reasoning, and emotion attributions: Relations with adolescents' reactive and proactive aggression. *Child Development, 80,* 1739-1755. doi:10.1111/j.1467-8624.2009.01365.x

Arsenio, W. F., & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child Development, 75*, 987-1002. doi:10.1111/j.1467-8624.2004.00720.x

Arsenio, W. F., & Lemerise, E. A. (2010). Introduction: An integrative approach to emotions, aggression, and morality. In W. F. Arsenio & E. A. Lemerise (Eds.), *Emotions, Aggression, and Morality in Children* (pp. 3-9). Washington, DC: American Psychological Association.

Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice Hall.

Berkowitz, L. (1963). *Aggression: A social psychological analysis*. New York, NY: McGraw-Hill.

Boldizar, J. P., Perry, D. G., & Perry, L. C. (1989). Outcome values and aggression. *Child Development, 60*, 571-579. doi:10.2307/1130723

Bosworth, K., Espelage, D., DuBay, T., Daytner, G., & Karageorge, K. (2000). Preliminary evaluation of a multimedia violence prevention program for adolescents. *American Journal of Health Behavior, 24*(4), 268-280. doi:10.5993/AJHB.24.4.3

Bowlby, J. (1969). *Attachment and loss* (Vol. I). New York, NY: Basic Books.

Bowlby, J. (1973). *Attachment and loss* (Vol. II). New York, NY: Basic Books.

Bowlby, J. (1980). *Attachment and loss* (Vol. III). New York, NY: Basic Books.

Bretherton, I (1985). Attachment theory: Retrospect and prospect. *Monographs of the Society for Research in Child Development, 50*, 3-35. doi:10.2307/3333824

Camodeca, M., & Goossens, F. A. (2005). Aggression, social cognitions, anger and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry, 46,* 186-197. doi:10.1111/j.1469-7610.2004.00347.x

Casey, R. J. (1996). Emotional competence in children with externalizing and internalizing disorders. In M. Lewis & M. W. Sullivan (Eds.), *Emotional development in atypical children* (pp. 161-183). Mahwah, NJ: Lawrence Erlbaum Associates.

Casey, R. J., & Schlosser, S. (1994). Emotional responses to peer praise in children with and without a diagnosed externalizing disorder. *Merrill-Palmer Quarterly, 40,* 60-81.

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorder: The FAST Track Program. *Development and Psychopathology, 4,* 509-528. doi:10.1017/S0954579400004855

Crick N. R., Casas, J. F., & Nelson, D.A. (2002). Toward a more comprehensive understanding of peer maltreatment: Studies of relational victimization. *Current Directions in Psychological Science, 11,* 98-101. doi:10.1111/1467-8721.00177

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74-101 doi:10.1037/0033-2909.115.1.74.

Crick, N. R., & Dodge, K. A. (1996). Social information processing mechanisms in reactive and proactive aggression. *Child Development, 67,* 993-1002. doi:10.2307/1131875

Crick, N. R., & Dodge, K. A. (1999). 'Superiority' is in the eye of the beholder: A comment on Sutton, Smith, and Swettenham. *Social Development, 8,* 128-131. doi:10.1111/1467-9507.00084

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710-722. doi:10.2307/1131945

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. doi:10.1111/1467-8624.00462

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69,* 1630-1639. doi:10.1111/j.1467-8624.1998.tb06181.x

Denham, S. A., & Burton, R. (1996). A social-emotional intervention for at-risk 4-year-olds. *Journal of School Psychology, 34,* 225-245. doi:10.1016/0022-4405(96)00013-1

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *Cognitive perspectives on children's social and behavioral development* (pp. 77-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dodge, K. A. (1991). Emotion and social information processing. In J. Garber & K. A. Dodge (Eds.), *The development of emotion regulation and dysregulation* (pp. 159-181). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511663963.009

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology, 18*, 791-814. doi:10.1017/S0954579406060391

Dodge, K.A., Bates, J.E., & Pettit, G.S. (1990). Mechanisms in the cycle of violence. *Science, 250,* 1678-1683. doi:10.1126/science.2270481

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. doi:10.1037/0022-3514.53.6.1146

Dodge, K. A., & Newman, J. P. (1981). Biased decision making processes in aggressive boys. *Journal of Abnormal Psychology, 90,* 375-379. doi:10.1037/0021-843X.90.4.375

Dodge, K. A., & Rabiner, D. L. (2004). Returning to roots: On social information processing and moral development. *Child Development, 75*, 1003-1008. doi:10.1111/j.1467-8624.2004.00721.x

Dollard, J., Miller, N. E., Doob, C. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press. doi:10.1037/10022-000

Dryfoos, J. G. (1990). *Adolescents at risk: Prevalence and prevention.* New York, NY: Oxford University Press.

D'Zurilla, T. J., Goldfried, M. R. (1971). Problem solving and behavior modification. *Journal of Abnormal Psychology*, *78,*107-126. doi:10.1037/h0031360

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53-88. doi:10.1146/annurev.ps.32.020181.000413

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

Eisenberg, N., Fabes, R. A., Nyman, M., Bernzweig, J., & Pinuelas, A. (1994). The relations of emotionality and regulation to children's anger-related reactions. *Child Development, 65*, 109-128. doi:10.2307/1131369

Elias, M. J. (1994). Capturing excellence in applied settings: A participant conceptualizer and praxis explicator role for community psychologists. *American Journal of Community Psychology*, *22,* 293-318. doi:10.1007/BF02506868

Farrington, D., Baldry, A.C., Kyvsgaard, B., & Ttofi, M. M. (2008). Effectiveness of programs to prevent school bullying. *Victims and offenders*, *2*, 183-204.

Feldman, E., & Dodge, K. A. (1987). Social information processing and sociometric status: Sex, age, and situational effects. *Journal of Abnormal Child Psychology, 15*, 211-227. doi:10.1007/BF00916350

Flavell, J. H., Botvin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children*. New York, NY: Wiley.

Fontaine, R. G. (2008). On-line social decision making and antisocial behavior: Some essential but neglected issues. *Clinical Psychology Review, 28*, 17-35. doi:10.1016/j.cpr.2007.09.004

Fontaine, R. G., & Dodge, K. A. (2006). Real-time decision making and aggressive behavior in youth: A heuristic model of response evaluation and decision (RED). *Aggressive Behavior*, *32*, 604-624. doi:10.1002/ab.20150

Forness, S. R., Serna, L. A., Nielsen, E., Lambros, K., Hale, M. J., & Kavale, K. A. (2000). A model for early detection and primary prevention of emotional or behavioral disorders. *Education and Treatment of Children, 23*, 325-345.

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. Y. (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*, 1045-1055. doi:10.1037/0022-006X.73.6.1045

Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders, 8*, 102-112. doi:10.1177/106342660000800206

Frey, K. S., Nolen, S. B., Van Schoiack-Edstrom, L.V.S., & Hirschstein, M. K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Applied Developmental Psychology, 26*, 171-200. doi:10.1016/j.appdev.2004.12.002

Gottman, J. M. (1986). Merging social cognition and social behavior. *Monographs of the Society for Research in Child Development, 51*(2), 81-85.

Gouze, K. R. (1987). Attention and social problem solving as correlates of aggression in preschool males. *Journal of Abnormal Child Psychology, 15*, 181-197. doi:10.1007/BF00916348

Grotpeter, J. K., & Crick, N. R. (1996). Relational aggression, overt aggression, and friendship. *Child Development, 67*, 2328-2338. doi:10.2307/1131626

Hartup, W. W. (1974). Aggression in childhood: Developmental perspectives. *American Psychologist*, *29*, 336-341. doi:10.1037/h0037622

Heckman, J. J. (2008). Schools, skills, and synapses. *Economic Inquiry, 46,* 289-324. doi:10.1111/j.1465-7295.2008.00163.x

Hubbard, J. A., Smithmyer, C. M., Ramsden, S. R., Parker, E. H., Flanagan, K. D., & Dearing, K. F. (2002). Observational, physiological, and self-report measures of children's anger: Relations to reactive versus proactive aggression. *Child Development, 73,* 1101-1118. doi:10.1111/1467-8624.00460

Huesmann, L. R. (1988). An information processing model for the development of aggression. *Aggressive Behavior, 14,* 13-24. doi:10.1002/1098-2337(1988)14:1<13::AID-AB2480140104>3.0.CO;2-J

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. I. Donnerstein (Eds.), *Human aggression: Theories, research, and implications for policy* (pp. 73-109). New York, NY: Academic Press.

Huesmann, L. R. & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72,* 408-419. doi:10.1037/0022-3514.72.2.408

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, UK: Cambridge University Press.

Ladd, G. W., & Mize, J. (1983). A cognitive-social learning model of social skill training. *Psychology Review, 90,* 127-157. doi:10.1037/0033-295X.90.2.127

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology, 34,* 715-724. doi:10.1007/s10802-006-9057-4

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71,* 107-118. doi:10.1111/1467-8624.00124

Lynch, K. B., Geller, S. R., & Schmidt, M. G. (2004). Multi-year evaluation of the effectiveness of a resilience-based prevention program for young children. *Journal of Primary Prevention, 24,* 335-353. doi:10.1023/B:JOPP.0000018052.12488.d1

Mandler, J. H. (1979). Categorical and schematic organization in memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 259-299). New York, NY: Academic Press.

Meyer, A. L., & Farrell, A. D. (1998). Social skills training to promote resilience in urban sixth-grade students: One product of an action research strategy to prevent youth violence in high-risk environments. *Education and Treatment of Children, 21,* 461-488.

Milich, R., & Dodge, K. A. (1984). Social information processing in child psychiatric populations. *Journal of Abnormal Child Psychology, 12*, 471-490. doi:10.1007/BF00910660

Murphy, B. C., & Eisenberg, N. (1997). Young children's emotionality: Regulation and social functioning and their responses when they are targets of a peer's anger. *Social Development, 6*, 18-36. doi:10.1111/j.1467-9507.1997.tb00092.x

Nelson, G., & Carson, P. (1988). Evaluation of a social problem-solving skills program for third- and fourth-grade students. *American Journal of Community Psychology, 16*, 79-99. doi:10.1007/BF00906073

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nucci, L., & Herman, S. (1982). Behavioral disordered children's conceptions of moral, conventional, and personal issues. *Journal of Abnormal Child Psychology, 10*, 411-426. doi:10.1007/BF00912330

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. doi:10.1111/1467-8624.00447

Pakaslahti, L. (2000). Children's and adolescents' aggressive behavior in context: The development and application of aggressive problem-solving strategies. *Aggression and Violent Behavior, 5*, 467-490. doi:10.1016/S1359-1789(98)00032-9

Parke, R. D. & Slaby, R. G. (1983). The development of aggression. In P. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), *Handbook of child psychology, Vol. 4: Socialization, personality, and social development* (4th ed., pp. 547-641). New York, NY: Wiley.

Patterson, G. (1986). Performance models for antisocial boys. *American Psychologist, 41*, 432-444. doi:10.1037/0003-066X.41.4.432

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology, 31*, 109-118. doi:10.1037/h0024436

Perry, D. G., Perry, L. C., & Rasmussen, P. (1986). Cognitive social learning mediators of aggression. *Child Development, 57*, 700-711. doi:10.2307/1130347

Pettit, G. S., Dodge, K. A., & Brown, M. M. (1988). Early family experience, social problem solving patterns, and children's social competence. *Child Development, 59*, 107-120. doi:10.2307/1130393

Pressley, M., EI-Dinary, P. B., Marks, M. B., Brown, R., & Stein, S. (1992). Good strategy instruction is motivating and interesting, In K. A. Remminger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 333-358). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rigby, K. (1996). *Bullying in schools and what to do about it.* Bristol, PA: Jessica Kingsley.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (pp. 37-86). Hillsdale, NJ: Erlbaum.

Rule, B. G. (1974). The hostile and instrumental functions of human aggression. In J. De Wit & W. W. Hartup (Eds.), *Determinants and origins of aggressive behaviors*. The Hague, Netherlands: Mouton.

Saarni, C. (1999). *The development of emotional competence.* New York, NY: Guilford Press.

Sawyer, M. G., Macmullin, C., Graetz, B., Said, J. A., Clark, J. J., & Baghurst, P. (1997). Social skills training for primary school children: A 1-year follow-up study. *Journal of Paediatrics and Child Health, 33*, 378-383. doi:10.1111/j.1440-1754.1997.tb01624.x

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structure.* Hillsdale, NJ: Lawrence Erlbaum Associates

Schultz, D., Ambike, A., Logie, A. K., Bohner, K. E., Stapleton, L.M., VanderWalde, H.,… Betkowski, J.A. (2010). Assessment of social information processing in early childhood: Development and initial validation of the Schultz test of emotion processing—preliminary version. *Journal of Abnormal Child Psychology*, *38*, 601-613. doi:10.1007/s10802-010-9390-5

Selman, R. L. (1971). Taking another's perspective: Role-taking development in early childhood. *Child Development, 42,* 1721-1734. doi:10.2307/1127580

Shantz, C. U. (1975). The development of social cognition. In E. M. Hetherington (Ed.), *Review of child development research* (Vol. 5, pp. 257-323). Chicago, IL: University of Chicago Press.

Shantz, C. U. (1983). Social cognition. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., p. 495-555). New York, NY: John Wiley & Sons.

Shapiro, J. P., Burgoon, J. D., Welker, C. J., & Clouch, J. (2002). Evaluation of the peacemakers program: School-based violence prevention for students in grades four through eight. *Psychology in the Schools*, *39*, 87-100. doi:10.1002/pits.10040

Shure, M. B., & Spivack, G. (1988). Interpersonal Cognitive Problem Solving. In R. H. Price, E. L. Cowen, R. P. Lorion, & J. Ramos-McKay (Eds.), *14 Ounces of Prevention: A Casebook for Practitioners* (pp. 69-82). Washington, DC: American Psychological Association.

Smith, M. (2001). Social and emotional competencies: Contributions to young African-American children's peer acceptance. *Early Education and Development, 12,* 49-72. doi:10.1207/s15566935eed1201_4

Spivack, G., & Shure, M. B. (1974). *Social adjustment of young children: A cognitive approach to solving real-life problems*. San Francisco, CA: Jossey-Bass.

Sroufe, L. A., Schork, E., Motti, F., Lawroski, N., & LaFreniere, P. (1984). The role of affect in social competence. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition, and behavior* (pp. 289-319). New York, NY: Cambridge University Press.

Strassberg, Z., & Dodge, K. A. (1987, November). *Focus of social attention among children varying in peer status.* Paper presented at the Annual Meeting of the Association for the Advancement of Behavior Therapy, Boston, MA.

Sutton, J., Smith, P. K., & Swettenham, J. (1999). Social cognition and bullying: Social inadequacy or skilled manipulation? *British Journal of Developmental Psychology, 17,* 435-450. doi:10.1348/026151099165384

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7,* 77-88. doi:10.1037/1528-3542.7.1.77

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. San Francisco, CA: Jossey-Bass

Wilson, S. J., & Lipsey, M. W. (2006). *The effects of school-based social information processing interventions on aggressive behavior. Part I: Universal programs.* Campbell Systematic Review 2006:5. doi:10.4073/csr.2006.5 Retrieved from http://campbellcollaboration.org/lib/download/68/

Winfrey, L. P. L., & Goldfried, M. R. (1986). Information processing and the human change process. In R. Ingram (Ed.), *Information processing approaches to clinical psychology* (pp. 241 -258). San Diego, CA: Academic Press.

Work, W. C., & Olsen, K. H. (1990). Evaluation of a revised fourth grade social problem solving curriculum: Empathy as a moderator of adjustive gain. *Journal of Primary Prevention, 11*, 143-157. doi:10.1007/BF01325280

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. doi:10.1037/0022-3514.77.1.150

Zillman, D. (1979). *Hostility and aggression*. Hillsdale, NJ: Lawrence Erlbaum Associates.

PAPER II

EVALUATING DOSAGE EFFECTS USING GENERALIZED PROPENSITY SCORE

Dosage analysis is an important but critically understudied area of social intervention research. Dosage analysis not only provides important data regarding the optimal amount of exposure to a social intervention but also enables researchers to untangle program effects from implementation effects. A primary challenge in conducting dosage analyses is the need to simultaneously balance multiple groups. A potential solution to this challenge is offered by generalized propensity score (GPS) methods, which are a relatively recent development within the family of propensity score statistical techniques. This paper first reviews issues encountered in early attempts to conduct dosage analyses, and then introduces the GPS methods used for conducting ordered, unordered, and continuous dosage analyses. The discussion is based around an example demonstrating the use of the GPS method with a continuous treatment variable. Challenges in applying GPS methods are discussed.

Overview

Social interventions are often delivered in varying quantity either as a planned element of a study or as a function of differential implementation by intervention agents (e.g., agency staff trained to deliver the program). Borrowing medical terminology, the varying amounts of social interventions are called *doses*. The dosage of social interventions can be measured in a variety of forms. The simplest forms of measuring dosage include tracking participants' direct exposure to the intervention content and recording minutes or hours of training classes (e.g., Guo & Fraser, 2009; Zhai et al., 2010), number of psychotherapy sessions (e.g., Howard, Kopta, Krause, & Orlinsky, 1986), number of mental health treatment sessions (e.g., Bickman, Andrade, & Lambert, 2002), or years of mental health consultation (e.g., Alkon, Ramler, & MacLennan, 2003). Dosage can also be measured as indirect exposure to program content, such as the number of media channels with family planning information (e.g., Jato et al., 1998). In addition, measures of dosage can involve simple calculations such as the ratio of attendance over classes offered (Miller & Dyk, 1991).

Because program effects typically vary across participants who have experienced different dosages of treatment, the evaluation of treatment effects at different dosage levels is critical to social intervention research, and is referred to as *dosage analysis* (e.g., Zhai et al., 2010) or *dose-response analysis* (e.g., Imbens, 2000). Assessing dosage effects in social intervention research is important to the arenas of practice and policy primarily for two reasons. First, the effects of social interventions are rarely a linear function of the amount of treatment or a case of "the more, the better". Therefore, determining the optimal dosage that produces the maximum beneficial results is of great interest to practitioners who are interested in helping clients achieve optimal outcomes, and to policy makers who are interested in ensuring program efficiency.

Second, the effects of social interventions are unavoidably influenced by factors related to differential implementation. When varying dosages of treatment are results of differential implementation, dosage analyses facilitate untangling theoretical program effects from implementation effects. Relying on reports of global program effects without accounting for implementation effects can lead to inappropriate conclusions, and the consequences are not trivial (Angrist, 2006; Fraser et al., 2011; Lochman, Boxmeyer, Powell, Roth, & Windle, 2006).

Although the importance of dosage analyses has long been recognized by social researchers (e.g., Howard et al., 986; Fraser et al., 2011; Peck, 2003), dosage analyses remain an understudied area (Zhai et al., 2010). One factor influencing the scarcity of research on dosage analysis is perhaps the strong emphasis that program funders place on intent-to-treat analyses (Fraser et al., 2011). In addition, methodological challenges have posed another factor that likely impeded research on dosage analyses. As previously mentioned, variations in dosage are often unplanned elements that result from differential implementation by intervention agents (e.g., only half of the treatment was delivered) or from noncompliance of participants (e.g., some participants missed half of the training classes). Because the dosage groups are not the product of random assignment to different treatment conditions, the dosage groups are not directly comparable and statistical measures have to be taken to account for overt selection bias (i.e., bias due to differences in observed covariates, Rosenbaum, 1991). Moreover, dosage analyses often involve comparing more than two groups. Balancing multiple groups simultaneously is particularly challenging and often requires advanced statistical technology rather than relying on conventional regression analyses or matching techniques used for two-group comparisons.

Dosage analyses have become more feasible given the recent development of generalized propensity score (GPS) methods that provide researchers with a viable means to balance multiple groups simultaneously. However, discussions of these methods have been largely confined to statisticians and economists, and the statistical approaches are relatively sophisticated (e.g., Hirano & Imbens, 2004; Imai & Van Dyk, 2004; Imbens, 2000; Joffe & Rosenbaum, 1999). This article aims to introduce social science researchers to the utility of GPS methods for intervention research. To fully explain the benefit of GPS methods, I first present a review of the methodological limitations encountered in early dosage analyses. Challenges in conducting dosage analysis using GPS and areas for future research are then discussed. Lastly, an example is provided by applying a GPS method to a dosage analysis when dosage varies continuously.

Limitations of Early Dosage Analyses

Early studies that estimated dosage effects were constrained by the inherent limitations of conventional regression methods and by focusing on two dosage groups (e.g., Andrade, Lambert, & Bickman, 2000). Regression analysis directly models the relationship between an outcome variable and confounding factors. This analytic approach estimates treatment effects by partitioning out effects due to observed confounders (Cochran, 1983; Cook & Campbell, 1979). When applied appropriately, regression analysis can yield estimates that approximate results from randomized experiments (Shadish, Clark, & Steiner, 2008). However, regression analysis has limitations. First, this method generally assumes that relationships between the potential confounders and the outcome of interest are linear.

Although interaction and nonlinear terms can be added to a regression model, the relationships between outcome and these transformed covariates remain fundamentally linear (Schafer & Kang, 2008). A regression model also assumes identical slope for confounders between treatment and control groups. The performance of a regression model can be sensitive to any departures from these assumptions.

Second, when the distributions of confounders in dosage groups differ substantially and the distributions have a relatively small overlap, then regression analysis involves a certain amount of extrapolation (i.e., comparing individuals who do not have comparable counterparts). Estimates involving extrapolation can be highly sensitive to functional form and prone to bias due to misspecification (Drake, 1993; Rubin, 1997). For a more detailed description of dangers of model-based extrapolation see King (2006) and King and Zeng (2006).

Third, regression model is limited by concerns about overfitting. When the number of potential confounders is large, it might be impossible to include all potential confounders, interaction terms, and non-linear terms in a regression model. Omitting any potential confounders makes the estimates inclined to bias (Orwin et al., 2003).

The limitations of regression model might account to a certain extent for the mixed findings from multiple studies that evaluate dosage effects of the Fort Bragg Demonstration, a mental health project for children. Using regression analysis, two studies found no effects (Andrade et al., 2000; Salzer, Bickman, & Lambert, 1999). However, findings from studies that used instrumental variable method (Foster, 2000) and propensity score method (Foster, 2003) consistently found positive effects of the program.

Another major limitation of early studies that evaluated dosage effects was the focus on the comparisons of only two dosage groups (e.g., Andrade et al., 2000; Foster, 2003; Lochman et al., 2006). In practice, multiple dosage groups often exist either as planned or as a result of differential implementation. When there are many dosage levels, then treating the dosage variable (e.g., minutes of training, number of psychotherapy sessions) as a continuous variable might be appropriate. Collapsing the dosage variable that takes on multiple values into two levels usually results in loss of information and leaves the effects of varying dosage hidden.

GPS: An Extension of Propensity Score to Multivalued Treatment Settings

GPS methods are a relatively recent development in the growing family of propensity score-based methods. GPS methods expand the application of propensity score methods from binary treatment settings (Rosenbaum & Rubin, 1983, 1984) to multivalued treatment settings (Imbens, 2000; Joffe & Rosenbaum, 1999; Lechner, 2001) and continuous treatment settings (e.g., Behrman, Cheng, & Todd, 2004; Hirano & Imbens, 2004; Imai & Van Dyk, 2004). GPS shares the key property of propensity score, that is, they are balancing scores (Joffe & Rosenbaum, 1999; Hirano & Imbens, 2004; Imbens, 2000). However, moving from binary treatment settings to multivalued treatment settings requires modifications to the definition of propensity score and to the assumption of unconfoundedness. The estimation of GPS also uses different procedures than those used for estimating propensity score with binary treatment.

Definition of GPS. In binary treatment settings, the propensity score is defined as "the conditional probability of exposure to a treatment given observed covariates" (Rosenbaum & Rubin, 1983, p.41), and can be denoted as $e(x) \equiv pr(T = 1 \mid X = x)$, where T

is the treatment, and X is a set of covariates. The GPS with multivalued treatment is defined as “the conditional probability of receiving a particular level of the treatment given the pretreatment variables” (Imbens, 2000, p. 708), and can be expressed as $r(t,x) \equiv pr(T = t \mid X = x)$. First coined by Imbens (2000), the term GPS was used for unordered treatment settings. The term has since been used to refer to propensity scores with nonbinary treatment settings (e.g., Imai & Van Dyk, 2004).

Fundamental assumptions. To enable drawing causal inferences, propensity score methods with binary treatment rely on two fundamental assumptions: the first assumption is the *stable unit treatment value assumption* (SUTVA, Rubin, 1978, 1980); and the second is the unconfoundedness assumption (Rubin, 1990). SUTVA states that a participant’s outcome is not affected by other participants’ treatment assignments. A major implication of this assumption is that no social interaction takes place among study participants. Applying GPS methods in estimating dosage effects requires the same SUTVA assumption.

Unconfoundedness refers to a situation in which treatment assignment is independent of the potential outcomes conditioning on observed covariates. To explain the unconfoundedness concept in practical terms means that adjusting for differences in a fixed set of covariates removes biases in comparisons between treated and control participants, thus allowing for a causal interpretation of the adjusted outcome differences. The critical implication of the unconfoundedness assumption is that there are no unobserved confounders. Using notation, the unconfoundedness assumption can be expressed as $T \perp Y(t) \mid X$, where $Y(t)$ is the potential outcome associated with each participant and each value of the treatment t (Rosenbaum & Rubin, 1983). The notation $A \perp B \mid C$ represents independence between variables A and B given an event C (Dawid, 1979). The unconfoundedness assumption has been referred to by different names such as the *strongly ignorable treatment assignment assumption* (Rosenbaum & Rubin, 1983), *exogeneity* (Imbens, 2003), *selection on observables* (Barnow, Cain, & Goldberger, 1980; Fitzgerald, Gottschalk, & Mofitt, 1998), or *conditional independence* (Ichina, Mealli, & Nannicini, 2008; Lechner, 1999).

It has been proved that if treatment assignment is unconfounded given the pre-treatment variables, then treatment assignment is unconfounded given the propensity score. The unconfoundedness given the propensity score implies that average outcomes can be estimated by conditioning solely on the propensity score. This unconfoundedness assumption is rather strong. When applied to multivalued treatment settings, this assumption requires the treatment T to be independent of the entire set of potential outcomes.

Imbens (2000) introduced a weak version of the unconfoundedness assumption and has proven that the weak version is sufficient to validate causal estimation of average outcomes. The weak unconfoundedness assumption requires conditional independence of each level of the treatment with its associated potential outcomes, rather than joint independence of all potential outcomes for all dose levels (Imbens, 2000). The weak unconfoundedness assumption can be denoted as $D(t) \perp Y(t) | X$, where $D(t)$ is the indicator of receiving a specific treatment level t, and takes on a value of either 1 or 0. Similar to the case in binary treatment settings, assuming treatment assignment is weakly unconfounded given pretreatment variables X, then treatment assignment is weakly unconfounded given GPS $r(t, x)$ (Imbens, 2000). The implication is that it is sufficient to solely adjust for GPS to remove biases associated with pretreatment variables.

Estimation procedure. Another major difference between propensity score methods and GPS methods is that the propensity scores and the GPS are estimated using different procedures. Logistic regression is the standard approach in estimating the propensity score with binary treatment (Rosenbaum & Rubin, 1984). In contrast, no single standard approach is used for estimating GPS in multivalued treatment settings. In fact, various methods are needed based on the characteristics of the treatment values. When treatment takes on multiple values, the values can be qualitatively distinct and without a logical ordering, such as medication versus mindfulness meditation for drug abuse. The multiple values of the treatment can also be ordered and discrete (e.g., dose of a drug) or continuous (e.g., length of social skills training). Researchers have developed methods for estimating GPS with each of the three types of treatment variables. The next section introduces three GPS estimation procedures and the application of the estimated GPS in estimating dosage effects.

Three GPS Methods

GPS Method for Ordered Doses

Joffe and Rosenbaum (1999) first extended propensity score-based methods to a multivalued treatment circumstance. These researchers proposed that under certain circumstances, a single scalar propensity score existed with multiple doses. An example of a situation meeting the above criteria would be when the dose is ordered and the conditional distribution of doses given covariates *X* can be accurately described by McCullagh's (1980) ordinal logit model. Although Joffe and Rosenbaum's idea was novel, their proposal was brief and did not provide the practical guidance that applied researchers needed.

The Joffe and Rosenbaum proposal was extended by Lu, Zanutto, Hornik, and Rosenbaum (2001), who applied the method to a dose-response analysis using a propensity score matching procedure. Lu et al. evaluated the dose effects of exposure to a media campaign on intentions for future drug use. Five doses were defined based on the amount of exposure, and the dosage analysis involved four steps. First, a single scalar propensity score was estimated using an ordered logistic regression model. Second, the distance between two participants was calculated. Unlike the distance in binary treatment, which measures only the difference in observed covariates, the distance between participants in the dose-effect study takes into account both the difference in covariates and the difference of participants' dose levels. The formula for calculating the distance is notated as

$$d = \frac{\left(ps_k - ps_{k'}\right)^2 + \varepsilon}{\left(t_k - t_{k'}\right)^2},$$

where ps_k and $ps_{k'}$ are the estimated propensity scores for participants k and k'; t_k and $t_{k'}$ are the dose values for the two participants, respectively; ε is a vanishingly small but strictly positive number. The ε has two functions: (a) If two participants have the same dose (i.e., $t_k - t_{k'} = 0$), the distance d is ∞ even if they have the same propensity score (i.e., $ps_k - ps_{k'} = 0$); (b) If ($ps_k - ps_{k'}$) = 0, ε assures that d decreases as ($t_k - t_{k'}$) increases. A distance calculated this way enables researchers to match pairs that are similar on covariates but dissimilar on dosage.

Third, a nonbipartite pair matching was conducted using the distance scores calculated in the second step. A matching with two disjoint groups (e.g., under binary treatment condition) is called a bipartite matching (Rosenbaum, 1989). Matching between dose groups uses nonbipartite matching that employs a different algorithm than the one used in bipartite matching (for a detailed explanation, see Lu et al., 2001). Finally, when balance is achieved, dose effects are estimated by averaging outcome differences across all matched pairs. The significance test uses a Wilcoxon-signed-rank test. The test is a nonparametric statistical hypothesis test, and can be used as an alternative to the paired Student's *t*-test when the population cannot be assumed to be normally distributed or the data is on the ordinal scale (Kiess, 2002). To be sure, although multiple doses are defined, a significant dose effect does not imply that a dose effect exists in any pair of doses. The dose effect is generalized across all the comparisons; the only implication of the dose effect is that, on average, more exposure has the potential to yield a better outcome.

Zanutto, Lu, and Hornik (2005) further extended the method with ordered dosages. In the first stage, Zanutto and colleagues estimated a single scalar propensity score in the same way as Lu et al. (2001) used ordered logistic regression. However, in the second stage, instead of using the nonbipartite pair matching techniques, Zanutto et al. employed a subclassification procedure based on an estimated GPS. If GPS values are adequately estimated within each stratum, then participants would be balanced across dose groups within strata. After balance is achieved, dose effects for each dose level are first estimated within each stratum and then estimated across all strata. In dose analysis, subclassification is easier to implement than matching because the analysis can be accomplished using standard statistical software, whereas nonbipartite matching requires the researcher to use a specialized code.

GPS Method for Unordered Doses

Imbens (2000) proposed a novel GPS approach that can be applied to unordered treatment. Imbens' approach estimates the probability of an individual receiving each of the multiple doses given observed covariates. Using this approach, an individual would have multiple propensity scores. Each propensity score corresponds to each treatment level.

Imbens was the first to label this application as the GPS method. The term GPS has been used since to refer to a propensity score that is generalized to nonbinary treatment settings, including the single scalar propensity score and the multiple propensity score (Imai & Van Dyk, 2004). Imben's approach generally involves two steps. In the first step, multiple propensity scores (i.e., GPS) are estimated using a multinomial logit model or multinomial probit model. In the second step, the researcher estimates the dose effect by adding the GPS directly as a covariate or by using inverse of the scores as weights in the outcome model. An example of using inverse of GPS as weights to estimate dosage effects can be found in Guo and Fraser (2010). Matching and stratification are not suitable with multiple propensity scores because the propensity scores for separate doses are different functions of covariates. Propensity scores of the same numeric value—but which represent different doses—are not equivalent substantively, and it is not possible to match individuals with the "same" propensity score in different doses (Imbens, 2000).

GPS Method for Continuous Doses

More recently, Hirano and Imbens (2004) developed a 4-step method to deal with continuous treatment. In step 1, the conditional distribution of the treatment (T) given covariates is estimated. It is assumed that the treatment or its transformation has a normal distribution conditional on the covariates:

$$g(T_i) \mid X_i \sim N\{(\beta_0 + \beta_1^{'} X_i), \sigma^2\},$$

where $g(T_i)$ is a transformation of the treatment variable that can satisfy the normality assumption about the treatment variable. Parameters β_0, β_1, and σ^2 are estimated by maximum likelihood. In step 2, the GPS is estimated by modeling the conditional density of the treatment given covariates and using a simple normal density function:

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2}[g(T_i) - (\hat{\beta}_0 + \hat{\beta}_1^{'} X_i)]^2\right)$$

where $\hat{\beta}_0$, $\hat{\beta}_1^{'}$, and $\hat{\sigma}^2$ are parameters estimated from the first step; $(\hat{\beta}_0 + \hat{\beta}_1^{'} X_i)$ is the conditional expectation of treatment. In step 3, the conditional expectation of the outcome is estimated as a flexible function of two scalar variables: the treatment (T_i) and the estimated GPS ($\hat{R}_i$). The model may include higher-order terms, interaction terms of the treatment variable, and the estimated GPS. When used with a quadratic approximation, the model can be written as:

$$E[Y_i \mid T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 \hat{R}_i + \alpha_4 \hat{R}_i^2 + \alpha_5 T_i \hat{R}_i .$$

The parameters are estimated by ordinary least squares. In step 4, the estimated parameters (from Step 3) are used to estimate the average potential outcome at each treatment level of interest (t):

$$E[Y(t)] = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha}_0 + \hat{\alpha}_1 \cdot t + \hat{\alpha}_2 \cdot t^2 + \hat{\alpha}_3 \cdot \hat{r}(t, X_i) + \hat{\alpha}_4 \cdot \hat{r}(t, X_i)^2 + \hat{\alpha}_5 \cdot t \cdot \hat{r}(t, X_i)\right),$$

where $\hat{r}(t, X_i)$ is the estimated GPS at treatment level t given X_i.

Challenges in Applying GPS Methods

Among the family of propensity score methods, GPS methods are relatively recent developments. Procedures for applying GPS methods are diverse and new approaches continue to emerge. Some procedures for applying the GPS are straightforward extensions of propensity score methods for binary treatment settings. However, applying the GPS also requires developing new statistical procedures, which presents many challenges. These challenges reside in identifying common support and testing balance when treatment takes on continuous values and assessing the plausibility of the unconfoundedness assumption. For applied researchers, a further challenge to using the GPS is posed by the lack of a statistical software package for conducting the GPS procedures, although most of the analytical procedures can be carried out separately using standard statistical software.

Assessing Common Support and Testing Balance

Identifying common support and assessing balance are two closely related issues that are relevant to propensity score methods. Strictly speaking, common support is the "overlap" of the multidimensional distribution of all relevant characteristics between groups before treatment. Similarly, balance is the match of two multidimensional distributions of all covariates of the treated and comparison groups (Stuart, 2010). The goal of any propensity score method is to construct groups that are balanced before treatment. The existence of sufficient common support is a precondition for achieving balance. Assessing common support and testing balance are two key procedures in any propensity score method.

Assessing common support. According to the strict definition of common support, a common support region should be identified by comparing multidimensional distributions of all covariates. However, this approach is not feasible and, therefore, alternatives that use lower-dimensional measures are needed. Similar to the propensity score with binary treatment, GPS offers an alternative. The GPS summarizes multidimensional characteristics of an individual into a single score. In binary treatment settings, the common support region can be identified as the propensity score region shared by the two groups under comparison (Stuart, 2010). Individuals outside the common support region are those who have extreme propensity scores and who do not have comparable counterparts in the opposite condition. Outliers with extreme propensity scores should be excluded from further analysis (Heckman, Ichimura, Petra, & Todd, 1997; Dehejia & Wahba, 1999).

Extending the approach from identifying common support in binary treatment settings to identifying common support in multivalued treatment settings is straightforward. When the multiple values of treatment have an inherent order, the GPS is a single scalar score estimated with ordered logistic regression. The common support is the GPS region that contains observations with all treatment levels (Zanutto et al., 2005). When the multiple values are qualitatively distinct and do not have a logical ordering, the GPS is estimated with the multinomial logit model (Imbens, 2000). Each individual has multiple propensity scores. Common support is inspected with GPS associated with each of the treatment levels separately. For GPS associated with a particular treatment level, the region of common support is the GPS region that contains observations for all treatment levels (Spreeuwenberg et al., 2010).

Identifying common support in continuous treatment settings is challenging because there are an infinite number of "treatment groups" and GPS to compare. One strategy to address this challenge involves a three-step process. First, the sample is divided into equal groups according to the treatment variable. Second, the GPS for the entire sample is estimated at the median or mean value of each treatment duration. Third, with each set of GPS, common support is assessed by comparing the GPS for the group with the treatment duration where GPS is estimated and the GPS for the rest of the sample. Individuals who have a GPS outside the common support regions are then excluded from the analytic sample. For examples, see Flores et al. (2010) and Kluve, Schneider, Uhlendorff, and Zhao (2012). It is important to note that this approach involves arbitrary decisions on the number of treatment groups into which the sample is divided and the treatment value at which GPS is estimated. For example, the sample can be divided into three groups or five groups according to the treatment variable. The treatment value chosen to estimate the GPS can be the median or the mean. Different choices are not only likely to result in different common support regions but also likely to yield different groups of individuals to be excluded from the analytic sample. Thus, an ongoing challenge for researchers remains the development of methods to assess common support with continuous treatment.

An issue closely related to identification of common support is the interpretation of the results. When a common support region is imposed and an analysis is restricted to a subsample, the interpretation of estimated treatment effects is conditioned. For example, the literature has suggested that propensity score weighting can be applied to estimate an average treatment effect for the population from which the individuals are sampled. However, when observations outside the common support region are excluded from analysis, researchers can no longer reliably estimate the average treatment effect of the population. The treatment effect applies only to individuals whose propensity scores fall within the common support region. An analysis of the characteristics of excluded cases compared with retained cases is often useful in determining the group for which the results apply (Crump et al., 2009).

Testing balance. Similar to propensity score methods in binary treatment settings, the essential value of GPS methods resides in the balancing property of the GPS. The use of GPS methods is valid only if balance can be improved after applying GPS. For propensity score methods with binary treatment, balance is the final criterion in appraising competing methods (Ho, Imai, King, & Stuart, 2007). Likewise, balance is also the final criterion for GPS methods with multivalued and continuous treatment. Consequently, reporting covariate balance before and after applying GPS should be a routine practice.

Balance is the similarity of two multidimensional distributions of all covariates of the treated and comparison groups (Stuart, 2010). Consequently, balance should be assessed by comparing the joint distributions of covariates in the treated and comparison groups. However, similar to the problem in identifying common support, practical strategies are not available for balance checking based on multidimensional distributions. Researchers have to find alternatives that use lower-dimensional measures. The most common practice is to compare marginal distributions of each covariate.

A commonly used approach to assess covariate balance in multivalued and continuous treatment settings is to regress each covariate on the treatment variable without and with conditioning on the estimated GPS (e.g., Kluve et al., 2012; Spreeuwenberg et al., 2010; Zanutto et al., 2005). For continuous covariates, the preferred choice is a linear regression model. For binary covariates, the researcher should use a logistic regression model. However, methods for including the estimated GPS vary across the three GPS methods. When the treatment variable takes on ordered values and the sample is stratified based on a single scalar score estimated with ordered logistic regression, balance should be assessed within each stratum. Zanutto and colleagues (2005) assessed balance within each quintile across the treatment levels using a regression model in which the dependent variable was the covariate, and the two independent variables were the treatment and the quintile of GPS. When the treatment variable is categorical and the GPS is estimated with the multinomial logit model, each individual will have multiple GPS values. Spreeuwenberg et al. (2010) assessed balance by regressing each covariate on the categorical treatment variable and the multiple sets of GPS.

Multiple methods have been introduced to test balance when treatment is continuous. Flores and colleagues (2010) used a gamma model with a log link for the treatment variable. The model included all covariates employed in the GPS model and the estimated GPS up to a cubic term (the unrestricted model). The unrestricted model was then compared with a restricted model that set the coefficients of all covariates to zero using a likelihood ratio test. Flores and colleagues' rationale was that if the GPS sufficiently balanced the covariates, then the covariates could be excluded from the model because the covariates would have little explanatory power conditional on the GPS. The researchers chose this method because they had a large number of covariates in addition to interaction terms and higher-order polynomials.

Kluve et al. (2012) used multiple methods for balance check. One method regressed each covariate on the treatment variable and the GPS. The GPS was evaluated at the 25th, the 50th and the 75th percentile of the treatment duration. If the GPS sufficiently balanced the covariates, then the treatment variable would be uncorrelated with the covariate. Another approach used by Kluve et al. (2012) was labeled as "blocking on the score." In this approach, the sample was divided into three groups at the 30th and 70th percentiles of the distribution of length of treatment. Within each group, the GPS was evaluated at the median of the treatment variable. Each group was then divided into five blocks by the quintiles of the GPS estimated at the median. For individuals whose GPS fell in the same quintile, differences in means of covariates were calculated between individuals whose treatment level belonged to a particular treatment level group and those whose treatment level was outside the particular treatment level group. The *t*-statistic of the differences in means between the particular treatment level group and all other groups was calculated using the weighted average over the five blocks in each treatment level group. The procedure was repeated for each treatment level group and for each covariate.

Assessing the Plausibility of the Weak Unconfoundedness Assumption

A causal interpretation of the estimated dosage effects is contingent on the plausibility of the weak unconfoundedness assumption. Similar to the strong unconfoundedness assumption for estimating treatment effects with binary treatment, the weak unconfoundedness assumption implies that groups are balanced on observed covariates and there are no unobserved confounders. Although balance on observed covariates can be evaluated, it is impossible to directly test for unobserved confounders. To make the unconfoundedness assumption plausible, researchers must identify and collect data on all speculated confounders. The identification of confounders requires sophisticated theory and cumulative evidence from empirical studies concerning relevant covariates.

Although the unconfoundedness assumption is not directly testable, methods have been developed to assess the assumption indirectly in cases in which treatment is binary. One approach focuses on estimating a causal effect that is known to equal zero. This approach can be applied when multiple control groups are available. If the researcher can assume the multiple control groups have similar distributions of observed covariates, then the researcher can expect to see a zero “average treatment effect” when making comparisons between the control groups. If the average treatment effect turns out to be nonzero, then the nonzero effect is attributable to unmeasured covariates omitted from the analysis. Under such a circumstance, unconfoundedness does not hold (Rosenbaum, 1987a; Heckman & Hotz, 1989; Heckman et al., 1997). Although the idea underlying this approach is intuitive, using this approach in practice is often not feasible because it requires multiple control groups.

Moreover, a second control group is useful only when that group can provide supplementary information about potential unobserved biases that the researcher thinks might be present (Rosenbaum, 1987b). This approach has not been extended to assess the unconfoundedness assumption when there are multiple treatment groups.

Another approach to testing the assumption of unconfoundness is *sensitivity analysis*. A sensitivity analysis "determines the magnitude of hidden bias that would need to be present to alter the conclusions of an observational study" (Rosenbaum, 2003, p. 2). Hidden bias is the bias that results from unobserved covariates. If an unreasonably strong assumption about hidden bias is required to alter the conclusions of a study, then bias is considered unlikely to exist. Thus, a causal conclusion becomes more defensible against the argument of confounding from unobserved covariates. Several different methods have been developed for conducting a sensitivity analysis in binary treatment settings (e.g., Brumback, Hernan, Haneuse, & Robins, 2004; Harada, 2012; Ichino et al, 2007; Lin, Psaty, & Kronmal, 1998; Pearson, 2003). These methods share a basic idea: to include a hypothetical unobserved covariate U in the analysis and assess the change in results under a range of assumptions about U (e.g., Bross, 1966, 1967; Cornfield et al., 1959; Imbens, 2003; Rosenbaum, 1987a, 2002, 2005; Rosenbaum & Rubin, 1983). However, these sensitivity analysis methods have not been extended to settings where treatment takes multiple values. Developing methods to conduct sensitivity analyses in multiple and continuous treatment settings remains a challenging area for future studies.

Availability of Software

Most of the procedures involved in applying GPS methods can be carried out using standard statistical software. However, nonbipartite matching requires the use of a specialized code. Although codes have been created for nonbipartite matching in C, FORTRAN, and R languages based on different algorithms, none of these codes are available in standard statistical software (Lu, Greevy, Xu, & Beck, 2011). They are only available from the developers. The C code is based on Gabow's (1973) algorithm, and can be downloaded from *htttp://elib.zib.de/pub/Packages/mathprog/matching/weighted/index.html*. The FORTRAN codes for the nonbipartite matching were created by Derigs (1988). Recently, Lu et al. (2011) created an R package based on Derigs's algorithm. It is free and can be downloaded from the CRAN website http://cran.r-project.org/web/packages/nbpMatching and the Vanderbilt Biostatistics website http://biostat.mc.vanderbilt.edu/NonbipartiteMatching. The later one is for those who are not familiar with R.

Bia and Mattei (2008) created a STATA package *doseresponse.ado* for the GPS method developed by Hirano and Imbens (2004). However, this package does not include codes for evaluating the common support region. Moreover, assuming all covariates are balanced after incorporating the GPS, the codes do not allow adding unbalanced covariates in the outcome model. In practice, it is not unusual for some covariates to remain unbalanced after numerous iterations of specifying the GPS model and testing balance. If this is the case, researchers need to write their own codes to include unbalanced covariates in the outcome model.

Application

This section provides an example of a dosage analysis for a setting with continuous treatment using the method introduced by Hirano and Imbens (2004). Dosage analysis in situations with continuous treatment is a relatively new development among GPS methods.

To the best of my knowledge, this GPS method has not been applied to assessing dosage effects in social work research. Assessing the effects of length or intensity of a treatment represents an important but understudied line of inquiry.

The data used in this analysis were obtained from a longitudinal study of *Making Choices* (Fraser et al., 2009), which is a social- and emotional-skills training program for elementary school children. The primary goals of the *Making Choices* program were to promote social competence and to reduce aggressive behavior in elementary school children. Participants were third-grade students in the 2004 and 2005 cohorts from 14 schools in North Carolina. The study used a cluster randomization design that first matched schools into pairs based on five key school-level characteristics. Then, schools within each pair were randomly assigned to either the treatment or the control condition. Students in the treatment condition received 28 *Making Choices* core lessons during their Grade 3 year, and 8 *Making Choices* follow-up or “booster shot” lessons in Grade 4 and Grade 5. In addition, teachers in the intervention condition received training and consultation on classroom behavior management and peer social dynamics.

Only participants assigned to the treatment condition were included in this dosage analysis. The exclusion of the control group is due to both substantive and statistical considerations. First, the primary goal of a dosage analysis is to identify optimal doses rather than to evaluate the overall effects of a treatment (for overall effects of *Making Choices*, see Fraser et al., 2009). Second, this dosage analysis treats the treatment variable as continuous and assumes a normal distribution of the observations. As such, including control participants with zero minutes of treatment would violate the normality assumption.

The sample consisted of 400 students from 30 classrooms (173 Black, 155 White, 30 Hispanic, 14 American Indian, and 28 other ethnicity). The majority of the sample was female (55%, n = 220), and 45% was male (n = 180). Baseline data were collected before students received any *Making Choices* lessons. New waves of data were collected each spring and fall over the course of the 3-year intervention study. The measures assessed program fidelity, including the minutes of *Making Choices* instruction delivered in each classroom. Preliminary analysis has shown the number of minutes of instruction varied widely among classrooms. Using the *minutes of instruction* variable as a measure of program dosage, this example evaluates dosage effects on children's social competence as measured at the end of Grade 3. The minutes of *Making Choices* instruction delivered in the third-grade year ranged from 268 to 2,340 minutes across 30 classrooms with a mean of 1,071.73 minutes, a median of 1,088 minutes, and a standard deviation of 385.67 minutes.

To investigate whether the treatment effects vary by the length of treatment received (i.e., minutes of instruction) using the GPS method proposed by Hirano and Imbens (2004), the analytical procedure followed five major steps:

Step 1. Estimating the conditional distribution of minutes given covariates X_i by maximum likelihood. In this analysis, a log transformation of minutes is applied to satisfy the normal distribution assumption about the treatment variable. The initial selection of included covariates was based on the theoretical and empirical association of each variable with the treatment (i.e., minutes of instruction) and the outcome (i.e., level of social competence). After iterations of specifying the model estimating the conditional distribution of minutes, estimating GPS, testing covariate balance, and respecifying the model estimating the conditional distribution of minutes, the final model included 26 linear terms and nine square terms. These covariates included variables measured at the student, classroom, and school levels.

Step 2. Estimating GPS by modeling the conditional density of the log transformation of minutes given covariates using a simple normal density function. The predicted value of treatment and standard deviation estimated in Step 1 were used in modeling the normal density function. The estimated GPS (i.e., the conditional density of the treatment given the covariates) ranged from 0.0004511 to 3.677995, with a mean of 2.722791 and a standard deviation of 0.97603.

Step 3. Identifying the common support region and testing balance. To identify the common support region, I followed the approach recommended by Flores et al. (2010). The sample was first divided into three subgroups of approximately equal size. The cut points were 1,030 minutes and 1,105 minutes. Three sets of GPS were then estimated at the median of each treatment interval. The common support region with respect to each treatment interval was obtained by comparing the GPS of participants belonging to the interval and those not belonging to the interval. The analysis sample was then limited to those participants whose GPS simultaneously occurred in the three common support regions. The overall GPS and the three sets of GPS estimated at the median of each treatment interval are presented in Table 2.1. The three common support regions defined by each set of the GPS are shown in Table 2.2.

Table 2.1
Overall GPS and the GPS Estimated at the Median of Each Treatment Interval

GPS		*N*	*M*	*SD*	Min.	Max.
Overall GPS		267	2.722791	.97603	.0004511	3.677995
GPS estimated at median of each treatment interval	GPS_1	267	.8512852	.933295	2.70e-29	3.607073
	GPS_2	267	2.159365	1.442738	2.10e-38	3.678463
	GPS_3	267	1.648196	1.648196	0	3.678609

Table 2.2
Common Support Region

Treatment interval with GPS estimate	Dosage group	GPS_1 Minimum	GPS_1 Maximum
≤1030 (Median = 966)	Minute≤1030	2.70e-29	3.607073
	Minute>1030	.0002477	3.584874
	Common support region 1: [.0002477, 3.584874]		
		GPS_2 Minimum	GPS_2 Maximum
1031-1105 (Median= = 1076)	1030<Minute≤1105	.0166222	3.678462
	Minute<1030 & Minute >1105	2.10e-38	3.678357
	Common support region 2: [.0166222, 3.678357]		
		GPS_3 Minimum	GPS_3 Maximum
>1105 (Median= = 1234)	Minute >1105	.0004511	3.678609
	Minute ≤1105	0	3.676587
	Common support region 3: [.0004511, 3.676587]		

To test covariate balance, I utilized a common approach by regressing each covariate on the treatment variable. A linear regression model was used to test the balance of continuous covariates. For binary covariates, I used a logistic regression model. Three sets of tests were conducted. The first test was done with the full sample before applying common support. After applying common support, the second and third sets of tests were conducted without and with conditioning on the estimated GPS.

Using a criterion of $p < 0.1$, 10 covariates were unbalanced before applying common support. The number of unbalanced covariates was reduced to four after applying common support (i.e., limiting the sample to children whose GPS fell in the common support region) but without accounting for the estimated GPS. There was no further reduction on the number of unbalanced covariates by accounting for the estimated GPS. It is worth noting that one covariate (i.e., hostile attribution) that was balanced initially became unbalanced after applying common support. Detailed information from the balance check is presented in Table 2.3. The table includes only the variables that were unbalanced (i.e., $p < 0.1$).

Table 2.3
Results of Balance Check

Covariate	Before Applying Common Support		After Applying Common Support			
			Without GPS		With GPS	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Demographic						
Hispanic	-.0019007	.001**	-.0003067	.929	-.0004128	.906
Student Report						
Cognitive concentration	.0006413	.004**	.0010989	.062^{+}	.0011028	.062^{+}
Academic achievement	.0007143	.025*	-.000438	.592	-.0003505	.669
Encoding	.0001771	<.001**	.0000328	.760	.0000307	.776
Hostile attribution	.0000315	.612	.0004526	.007**	.0004628	.006**
Teacher Report						
WOTD[a] *(s0_b4_wk?)*	.0012526	<.001**	.0080302	<.001**	.0079101	.<001**
PSS[b]	.0004057	.011*	.0004651	.335	.0004264	.378
Professional interest	.000818	<.001**	.0005689	.017*	.0005853	.014*
School Report						
PFRL[c]	-.0000853	.004**	.00002	.828	.0000238	.798
Adequate yearly progress	.0113292	<.001**	-.0004432	.847	-.0006342	.783
Income to poverty ratio	.078094	<.001**	.0922802	.0157	.0914177	.163

Note. a WOTD = Weeks devoted to tolerance and diversity activities. *b* PSS = Perception of student support. *c* PFRL = Percentage of students receiving free or reduced lunch.
** $p < .01$, * $p < .05$, $^{+}p < .10$, two-tailed

Step 4. Estimating the conditional expectation of the outcome. In Hirano and Imbens' (2004) approach, the conditional expectation of the outcome was estimated as a function of the treatment variable and the estimated GPS. Because the current analysis included unbalanced covariates, I followed the approach used in Abadie and Imbens (2002) and Lechner and Melly (2010) to include the unbalanced covariates in the regression model that estimates the conditional expectation of the outcome (i.e., the dose-response model). To control for rater effects, change score of social competence within Grade 3 was used as dependent variable. The results are presented in Table 2.4.

Table 2.4
Dose Response Function for Social Competence

Independent Variable	Social Competence (n = 210)	
	Coef.[a]	p value
Minute	-.005	.001**
GPS	-1.979	.003**
Minute-GPS	.002	.001**
Cognitive Concentration	-.135	.003**
Hostile Attribution	-.258	.101
WOTD	-.037	.338
Professional Interest	.363	.002**
Constant	4.178	0.023*

Note. WOTD = Weeks devoted to tolerance and diversity activities.
** $p < .01$, * $p < .05$, $^{+}$ $p < .10$, two-tailed

Step 5. Estimating the average potential outcome for each minute level of interest. The estimation was done by averaging the conditional expectation over the estimated GPS at the particular minute level of interest using the coefficients estimated in Step 4. Ten treatment levels were chosen, which included the lowest treatment level and the treatment levels that included approximately 10% to 100% of participants. The results are reported in Table 2.5.

Table 2.5
Average Dosage Effects

Intervention dose (by minutes of instruction)	Average outcome at each treatment dose
	Social competence (+)[a]
906	-.6316
945	.0658
1030	.0502
1068	.1332
1088	.0652
1105	.1363
1151	.2624
1234	.1372
1292	.3851
1380	-.2943

Note. [a] hypothetical sign

The results indicate that the intervention had dosage effects on social competence of children. However, the relationship between the increases of curriculum dose (i.e., minutes of *Making Choices* instruction) and the improvement of social competence is not strictly linear. Children who had similar dosages (e.g., 1234 vs. 1292) experienced substantially different treatment effects (i.e., 1372 vs. 3851). Moreover, the average dosage effects at the highest level (i.e., minute = 1,380) were negative. These seemly counterintuitive findings might be attributable to the quality of training. Many factors can affect the quality of training, including teacher characteristics (e.g., personality, teaching style) and classroom involvement (e.g., interest, investment). However, the data did not capture information on the quality of training and, therefore, the reasons for the decline in treatment effect remain unclear and pose an important topic for future study.

To adequately interpret this dosage analysis, several limitation of the analysis must also be considered. First, the method for assessing regions of common support involved arbitrary and subjective decisions. I chose to divide the sample into three subgroups based on length of *Making Choices* instruction and to estimate GPS at the median of each treatment interval. However, the sample could have been divided into quintiles or more subgroups, and the GPS could have been estimated at the mean value of each treatment interval. Second, the data used have a nested structure (i.e., students are nested within classrooms, and classrooms are nested within schools) that was not accounted for in the dosage analysis. For the social competence outcome, the intraclass correlation was .12 at school level and .25 at classroom level. By not accounting for nested data, the intraclass correlation might result in an underestimated standard error of the estimates and reduced power. Therefore, the *p*-value may be inflated, although the effect size would be unaffected by clustering. An ongoing challenge in conducting dosage analysis with continuous treatment will be the development of improved methods that will identify regions of common support more elegantly and objectively than current methods and will account for data with a nested structure.

Conclusions

Dosage analysis is an important line of inquiry. Findings from dosage analyses provide crucial information regarding optimal exposure (or doses) to an intervention. Policy decisions are often constrained by evaluation studies that report contradictory program findings (e.g., Malti, Ribeaud, & Eisner, 2011). One important explanatory factor for the contradictory program findings is varying implementation. In such situations, findings from dosage analyses facilitate untangling program effects from effects due to variation in implementation. Although the importance of dosage analysis in social science was recognized by researchers decades ago (Howard, Kopta, Krause, & Orlinsky, 1986), such analyses remain an understudied area. Recently, researchers have called for attention to the consequence of conflating program effects with implementation effects and reemphasized the importance of dosage analysis (Fraser et al., 2011)

Conducting a dosage analysis is a challenging pursuit for applied researchers. In part, the challenge in dosage analysis stems from the required task of simultaneously balancing multiple groups, a task that is typically beyond the capacity of conventional regression methods. To fill this gap, researchers have developed GPS methods as alternatives that provide viable means for balancing multiple groups. GPS methods are recent development in the family of propensity score methods. Similar to other statistical methods for making causal inferences, the successful application of GPS methods is contingent on the plausibility of some assumptions. GPS methods require a weak version of the unconfoundedness assumption. The key implication of this unconfoundedness assumption is the absence of unmeasured confounders. Thus, a successful use of GPS methods requires prudence in identifying and measuring confounders before embarking on the analysis. A crucial step in the analysis stage involves scrutinizing and specifying the GPS model through an iterative process of refining the model, identifying the common support region, and testing balance. Assessing common support and checking balance in situations with continuous treatment are challenging. Developing more elegant and less subjective approaches to identifying common support and checking balance represent an important area for future research.

REFERENCES: PAPER II

Abadie, A., & Imbens, G. (2002). Simple and bias-corrected matching estimators for average treatment effects. Retrieved from http://www.nber.org/papers/t0283.pdf?new_window=1

Andrade, A., Lambert, E. W., & Bickman, L. (2000). Dose effect in child psychotherapy: Outcomes associated with negligible treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(2), 161-168. doi:10.1097/00004583-200002000-00014

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why, and how. *Journal of Experimental Criminology, 2*, 23-44. doi:10.1007/s11292-005-5126-x

Alkon, A., Ramler, M., & MacLennan, K. (2003). Evaluation of mental health consultation in child care centers. *Early Childhood Education Journal, 31*(2), 91-99. doi:10.1023/B:ECEJ.0000005307.00142.3c

Behrman, J. R., Cheng, Y. M., & Todd, P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics, 86*(1),108-132. doi:10.1162/003465304323023714

Bickman, L., Andrade, A. R., & Lambert, E. W. (2002). Dose response in child and adolescent mental health services. *Mental Health Services Research, 4*(2), 57-70. doi:10.1023/A:1015210332175

Barnow, B. S, Cain, G. G., & Goldberger, A. S (1980). Issues in the analysis of selectivity bias. *Evaluation Studies, 5*, 42-59. Retrieved from http://www.eric.ed.gov/PDFS/ED202979.pdf

Bia, M., & Mattei, A. (2008). A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *Stata Journal, 8*(3), 354-373. Retrieved from http://www.stata-journal.com/article.html?article=st0150

Bross, I. D. J. (1966). Spurious effects from an extraneous variable. *Journal of Chronic Disease, 19*, 637-647. doi:10.1016/0021-9681(66)90062-2

Bross, I. D. J. (1967). Pertinency of an extraneous variable. *Journal of Chronic Disease, 20,* 487-495. doi:10.1016/0021-9681(67)90080-X

Brumback, B. A., Herman, M. A., Haneuse, S.J.P.A., & Robins, J. (2004). Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine, 23*, 749-764. doi:10.1002/sim.1657

Cochran, W. G. (1983). *Planning and analysis of observational studies.* New York, NY: Wiley. doi:10.1002/9780470316542

Cook, T., & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston, MA: Houghton Mifflin.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, *22*, 137-203. doi:10.1093/ije/dyp289

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika, 96*(1), 187-199. doi:10.1093/biomet/asn055

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B 41*(1), 1-31. Retrieved from http://www.jstor.org/stable/2984718

Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 94*, 1053-1062. Retrieved from http://www.jstor.org/stable/2669919

Derigs, U. (1988). Solving nonbipartite matching problems via shortest path techniques. *Annals of Operations Research, 13*(1), 225-261. doi:10.1007/BF02288324

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics, 49*, 1231-1236. doi:10.2307/2532266

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *Journal of Human Resources, 33*, 251-299. doi:10.2307/146433

Flores, C. A., Flores-Lagunes, A., Gonzalez, A., & Neumann, T. C. (2010). *Estimating the effects of length of exposure to instruction in a training program: The case of Job Corps.* Retrieved from http://moya.bus.miami.edu/~cflores/research_papers/FNGF-DRF-031610.pdf

Foster, E. M. (2000). Is more better than less? An analysis of children's mental health services. *Health Services Research*, *35*(5), 1135-1158. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089167/pdf/hsresearch00014-0080.pdf

Foster, E. M. (2003). Propensity score matching an illustrative analysis of dose response. *Medical Care, 41*, 1183-1192. doi:10.1097/01.MLR.0000089629.62884.22

Fraser, M. W., Guo, S. Y., Ellis, A. R., Day, S. H., Li., J. L., & Wike, T. L. (2009). *Social and character development in elementary school: Effects from a controlled trial.* Unpublished manuscript. School of Social Work, University of North Carolina, Chapel Hill, NC. Retrieved from http://sowkweb.usc.edu/download/research/social-and-character-development-elementary-school.pdf

Fraser, M. W., Guo, S. Y., Ellis, A., R., Thompson, A. M., Wike, T. L., & Li., J. L. (2011). Outcome studies of social, behavioral, and educational interventions: Emerging issues and challenges. *Research on Social Work Practice, 21*, 619-635. doi:10.1177/1049731511406136

Gabow, H. N. (1973). *Implementation of algorithms for maximum matching on nonbipartite graphs.* Unpublished doctoral dissertation. Stanford University. Retrieved from http://search.proquest.com/docview/287948024

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage.

Harada, M. (2012). *Generalized sensitivity analysis.* Retrieved from http://home.uchicago.edu/~masa/docs2/research.html#gsapaper

Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of Manpower Training**.** *Journal of the American Statistical Association, 84*, 862-874. Retrieved from http://www.jstor.org/stable/2290059

Heckman, J. J., Ichimura, H., Petra E., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies, 64*(4), 605-654. doi:10.2307/2971733

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp.73-84). New York, NY: John Wiley.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15,* 199-236. doi:10.1093/pan/mpl013

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist*, *41*(2), 159-164. doi:10.1037/0003-066X.41.2.159

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics, 23*(3), 305-327. doi:10.1002/jae.998

Imai, K., & Van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, *99*, 854-866. doi:10.1198/016214504000001187

Imbens, G. W. (2000). The role of propensity score in estimating dose-response functions. *Biometrika*, *87*, 706-710. doi:10.1093/biomet/87.3.706

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review, 93*(2), 126-132. doi:10.1257/000282803321946921

Jato, M. N., Simbakalia, C., Tarasevich, J. M., Awasum., D. N., Kihinga, C.N.B., & Ngirwamungu, E. (1999). The impact of multimedia family planning promotion on the contraceptive behavior of woman in Tanzania. *International Family Planning Perspectives, 25*(2), 60-67. doi:10.2307/2991943

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology, 150,* 327-333. doi:10.1093/oxfordjournals.aje.a010011

Kiess, H. O. (2002). *Statistical concepts for the behavioral sciences* (3rd ed.). Boston, MA: Allyn & Bacon.

King, G., & Zeng, L. (2006). The dangers of extreme counterfactuals. *Political Analysis, 14,* 131-159. doi:10.1093/pan/mpj004

King, G., & Zeng, L. (2007). When can history be our guide? The pitfalls of counterfactual inference. *International Studies Quarterly, 51*, 183-210. doi:10.1111/j.1468-2478.2007.00445.x

Kluve, J., Schneider, H., Uhlendorff, A., & Zhao, Z. (2012). Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society, 175*, 587-617. doi:10.1111/j.1467-985X.2011.01000.x

Lechner, M. (1999). *Identification and estimation of causal effects of multiple treatments under the conditional independence assumption*. Retrieved from http://ideas.repec.org/p/iza/izadps/dp91.html

Lin, D. Y., Psaty, B. M., & Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics, 54*, 948-963. doi:10.2307/2533848

Lochman, J. E., Boxmeyer, C., Powell, N., Roth, D. L., & Windle, M. (2006). Masked intervention effects: Analytic methods for addressing low dosage of intervention. doi:10.1002/ev.184

Lu, B., Greevy, R., Xu, X., & Beck, C. (2011). Optimal nonbipartite matching and its statistical applications. *American Statistician, 65*(1), 21-30. doi:10.1198/tast.2011.08294

Lu, B., Zanutto, E., Hornik, R., & Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association, 96,* 1245-1253. doi:10.1198/016214501753381896

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology, 40*, 677-692. doi:10.1080/15374416.2011.597084

McCullagh P. (1980). Regression models for ordinal data. *Journal of Royal Statistical Society, Series B, 42*, 109-142. Retrieved from http://people.csail.mit.edu/jrennie/papers/other/mccullagh-ordinal-80.pdf

Miller, B. C., Dyk, P. H. (1991). Community of caring effects on adolescent mothers: A program evaluation case study. *Family Relations, 40*, 386-395. doi:10.2307/584895

Orwin, R., Hornik, R., Judkins, D., Zador, P., Sridharan, S., & Baskin, R. (2003). *Innovative design and analysis strategies in the evaluation of the national youth anti-drug media campaign: Propensity scores and counterfactual projection weights in a national probability survey*. Retrieved from the Federal Committee on Statistical Methodology website. Available at http://www.fcsm.gov/03papers/Orwin.pdf

Pearson, R. K. (1998). *Generalized sensitivity analysis: A framework for evaluating data analysis results.* Retrieved from http://www.siam.org/proceedings/datamining/2003/dm03_20PearsonR.pdf

Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation*, *24*(2), 157-187.

Rosenbaum, P. R. (1987a). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, *74*(1), 13-26. doi:10.1093/biomet/74.1.13

Rosenbaum, P. R. (1987b). The role of a second control group in an observational study. *Statistical Science, 2*(3), 292-306. doi:10.1214/ss/1177013232

Rosenbaum, P. R. (1989). Optimal matching in observational studies. *Journal of the American Statistical Association, 84*, 1024-1032. Retrieved from http://www.jstor.org/stable/2290079

Rosenbaum, P. R. (1991). Discussing hidden bias in observational studies. *Annals of Internal Medicine, 115*, 901-905. Retrieved from http://annals.org/article.aspx?volume=115&page=901

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York, NY: Springer.

Rosenbaum, P. R. (2003). Does a dose-response relationship reduce sensitivity to hidden bias? *Biostatistics* , *4*(1), 1-10. doi:10.1093/biostatistics/4.1.1

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In B. S. Everitt & D. C. Howell (Eds.-in-Chief), *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 1809-1814). Hoboken, NJ: John Wiley & Sons. doi:10.1002/0470013192.bsa606

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55. doi:10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516-524. Retrieved from http://personal.psc.isr.umich.edu/yuxie-web/files/soc710/Rosenbaum-Rubin1984.pdf

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, *6*(1), 34-58. doi:10.1214/aos/1176344064

Rubin, D. B. (1980). Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by Basu. *Journal of the American Statistical Association, 75*(371), 591-593. doi:10.2307/2287653

Rubin, D. B. (1990). Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science, 5*(4), 472-480. doi:10.1214/ss/1177012032 . Retrieved from http://projecteuclid.org/euclid.ss/1177012032

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, 757-763. Retrieved from http://annals.org/article.aspx?volume=127&page=757

Salzer, M. S., Bickman, L., & Lambert, E. W. (1999). Dose–effect relationship in children's psychotherapy services. *Journal of Clinical and Consulting Psychology, 66*(2), 270-279. doi:10.1037/0022-006X.67.2.228

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279-313. doi:10.1037/a0014268

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*, 1334-1356. doi:10.1198/016214508000000733

Spreeuwenberg, M. D., Bartak, A., Croon, M. A., Hagenaars, J. A., Busschbach, J. J. V., Andrea, H., … Stijnen, H. (2010). The multiple propensity score as control for bias in the comparison of more than two treatment arms: An introduction from a case study in mental health. *Medical Care, 48*, 166-174. doi:10.1097/MLR.0b013e3181c1328f

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1-2. doi:10.1214/09-STS313

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics, 20*, 59-73. doi:10.3102/10769986030001059

Zhai, F. H., Raver, C. C., Jones, S. M., Li-Grining, C. P., Pressler, E., & Gao, Q. (2010). Dosage effects on school readiness: Evidence from a randomized classroom-based intervention. *Social Service Review, 84*, 615-655. doi:10.1086/657988

PAPER III

DISENTANGLING THE INTERVENTION EFFECTS OF A SIP-BASED SOCIAL-SKILLS TRAINING PROGRAM: DOSAGE ANALYSIS IN A CONTROLLED TRIAL OF THE *MAKING CHOICES* PROGRAM

Social-emotional skills training programs are ubiquitous in American public schools; however, evaluation studies of these programs have produced mixed findings. In particular, discrepancies are often found between results of small-scale efficacy studies and large-scale effectiveness trials, or between findings from studies led by program developers and studies led by independent investigators. Using data from a controlled trial of the *Making Choices* program, this article presents a dosage analysis that offers one method to investigate factors that might contribute to such discrepancies. A critical challenge in conducting dosage analysis is that dosage groups are often formed by nonrandom processes, including differential implementation. Groups with differing exposure to interventions may not be balanced. To address challenges in analyzing unbalanced data, this analysis used a generalized propensity score-based method with a continuous treatment variable. Dosage effects for *Making Choices* were evaluated for 8 key outcomes at the end of Grade 3 and Grade 4. Findings indicate that the intervention had dosage effects on social competence and emotional regulation at the end of Grade 3. In addition, the findings suggest that the qualitative aspects of implementation (e.g., the level of student engagement, teacher-student relationship) are an important area for future investigation.

Overview

Over the past few decades, researchers have established an association between early, persistent aggressive behavior and negative developmental outcomes, including poor academic performance, school drop-out, peer rejection, adolescent delinquency, and drug use (e.g., Bierman & Wargo, 1995; Coie & Kupersmidt, 1983; Crick & Grotpeter, 1995; Dodge & Pettit, 2003; Moffitt & Caspi, 2001; Prinstein & La Greca, 2004). Aggressive behavior in childhood also predicts adulthood outcomes such as unemployment, partner abuse, and violence (e.g., Farrington, 1998; Fraser, 1996; Odgers et al., 2008; Patterson, 2002). The detrimental impact of childhood aggressive behavior on life-course outcomes has led to a substantial number of studies dedicated to identifying developmental risk and protective factors and specifying strategies to disrupt links between early conduct problems and later maladjustment (e.g., Baldry & Farrington, 2005; Bollmer et al., 2005; Fraser, Kirby, & Smokowski, 2004; Jenson & Howard, 1999, 2001; Williams, Ayers, Van Dorn, & Arthur, 2004).

Children grow in a context defined by a variety of biopsychosocial factors. The development of childhood aggressive behavior has been found to be a complex function of various risk factors that reside in child, family, neighborhood, and school characteristics. These risk factors include difficult child temperament, poor social cognitive skills (Crick & Dodge, 1994), poverty (Brooks-Gunn & Duncan, 1997), harsh parenting (Eddy, Leve, & Fagot, 2001; Henry, Tolan, & Gorman-Smith, 2001; Wasserman & Seracini, 2001), family stress or adversity (Farrington, 1998; Fraser, 1996; Hanish & Guerra, 2002); peer rejection (Crick & Grotpeter, 1995; Guerra, Asher, & DeRosier, 2004); school violence, and neighborhood disorganization (Brendgen et al., 2008; Dodge & Pettit, 2003). Among these risk factors, the lack of social, cognitive and emotional regulation skills has been found to be an important mediator of aggressive behavior (Bandura, 1989; Dodge, 1980, 2006; Huesmann, 1988; Lengua, 2003; Zins, Weissberg, Wang, & Walberg, 2004). Moreover, recent findings have suggested that the social-emotional skills developed in childhood are as powerful as academic achievement in predicting adulthood socioeconomic success (Borghans et al., 2008; Heckman, Stixrud, & Urzua, 2006; Heckman & Kautz, 2012).

The importance of social-emotional skills to life course outcomes has led to the design and implementation of many school-based social-emotional skills training programs that aim to promote positive social development and decrease aggressive behavior (Wilson & Lipsey, 2007). Although these programs vary in their theoretical foundations and design features (e.g., program components, activities to carry out the theory), systematic reviews of evaluation studies of these programs have suggested that the majority of universal school-based social-emotional skills training programs were effective (Farrington & Welsh, 2003; Hahn et al., 2007; Payton et al., 2008; Wilson & Lipsey, 2003, 2006, 2007). However, a number of studies have reported mixed or even negative effects (e.g., Conduct Problems Prevention Research Group [CPPRG], 1999; Flannery et al., 2003; Grossman et al., 1997; Malti, Ribeaud, & Eisner, 2011; Merrell, Gueldner, Ross, & Isava, 2008; Park-Higgerson, Perumean-Chaney, Bartolucci, Grimley, & Singh, 2008; Multisite Violence Prevention Project, 2009). Moreover, even those programs shown to be effective in previous rigorous evaluation studies often showed much less or no effects when evaluated independent of the program developers (Eisner, 2009; Malti, Ribeaud, & Eisner, 2011).

Although the substantial and cumulating evidence supporting the efficacy and effectiveness of social-emotional skills training should lead to broader endorsement of these programs by researchers, practitioners, and policy makers, the mixed findings and the discrepancy of findings between program developers and non-developers raise the importance of investigating the effects of varying implementation.

The observed effects of a program are produced through a complex process on which a variety of factors can impinge. These factors can be roughly put under two categories: design and implementation. The key factors in designing a program are a solid theoretical foundation that provides a theory of change and activities that are both logically and linearly related to malleable mediators specified in the theory of change. Implementation is the process of carrying out a series of actions or activities intended to produce desired effects (Fraser, 2004). Assuming a well-designed program, variation in program effects from different evaluation studies might be largely the result of variation in implementation (i.e., what activities were used and how those activities were carried out).

A good implementation may be defined both quantitatively and quantitatively. For example, a school-based social-emotional skills training might include three dimensions related to implementation: (a) the quantity and quality of training and supervision for the classroom teachers who will deliver the intervention content; (b) the amount of student exposure to the training content (e.g., number of lessons taught or minutes of training delivered); and (c) the quality of classroom training (e.g., the level of student engagement, teacher-student relationship, teacher skill in presenting content).

Universal school-based social skills training programs typically consist of a standardized curriculum of sequential lessons. An important implementation variable is the length of participants' actual exposure to the training (i.e., *dosage*). The dosage received by each participant might vary widely for many reasons (e.g., school absences, student participation in other in-school activities). Previous studies have concluded that interventions need to have duration and intensity to show effects (Weissberg & Elias, 1993). Thus, it might be expected that varying the intervention dosage would result in different treatment effects. In program evaluation, it is increasingly important to go beyond the estimation of overall treatment effects and to further examine if and in what ways responses vary by the length of treatment exposure (i.e., dosage). Dosage analysis represents an emerging line of inquiry with the potential to help untangle core factors that influence treatment effects.

The purpose of this study was to evaluate the dosage effects of *The Competence Support Project*, which is an elementary school social and character development program. A primary task in conducting dosage analyses is controlling for selection bias through statistical means. Controlling for selection bias is foundational in dosage analyses because groups with varying program dosages are often not formed through randomization but rather they are formed as a result of variation in program implementation. One approach to controlling for potential selection bias is the use of propensity score-based methods. The current analysis used a generalized propensity score-based method introduced by Hirano and Imbens (2004) for investigating dosage effects with continuous treatment measures.

The Competence Support Project

The *Competence Support Project* (*CSP*) was one of seven social and character development (SACD) programs selected by the U.S. Department of Education Institute of Education Sciences and the Centers for Disease Control and Prevention to participate in a national evaluation of SACD programs. SACD programs have been widely implemented in American public schools. These programs share the goals of fostering the academic achievement and behavioral adjustment of elementary school-aged children. However, the various programs in the SACD study used different combinations of school and classroom activities that stemmed from different theoretical frameworks. The *CSP* intervention had three components: competence-enhancement behavior management (CEBM), social dynamics consultation, and the *Making Choices* social skills curriculum.

CEBM and social dynamics consultation involved teacher training and consultation on strategies for classroom behavior management. These strategies focused on rewarding appropriate behavior and providing logical consequences for inappropriate behavior (Fraser et al., 2009). These *CSP* components were implemented to regulate children's behavior and to improve the classroom environment by encouraging students to use social skills learned from the *Making Choices* curriculum.

The *Making Choices* curriculum was the core element of *CSP*. The *Making Choices* lessons were designed to promote children's social competence and reduce aggressive behavior by strengthening skills in processing social information and regulating emotions (Fraser, Nash, Galinsky, & Darwin, 200; Fraser et al., 2005; Nash, Fraser, Galinsky, & Kupper, 2003). *Making Choices* was primarily based on social information-processing (SIP) theory (Crick & Dodge, 1994). The SIP theory posits that a child's response to a social situation is formulated through cognitive processing, which occurs in a series of five overlapping steps that precede behavioral responses. The five cognitive steps include,

- Step 1, encoding of external and internal cues;
- Step 2, interpretation and cognitive representation of those cues;
- Step 3, clarification and selection of a goal;
- Step 4, response access or construction; and
- Step 5, response decision.

Further, the child's emotions are an integral element of each SIP step "in that emotion is the energy level that drives, organizes, amplifies, and attenuates cognitive activity and in turn is the experience and expression of this activity" (Dodge, 1991, p. 159; for reviews, see Crick & Dodge, 1994, 1996; Dodge, 2006).

Crick and Dodge's (1994) SIP model emerged from a research tradition based on the premise that social cognitions are the mechanisms leading to social behaviors. Earlier work using this approach focused on global (or *off-line*) cognitive constructs such as perspective taking, role taking, and referential communication (e.g., Flavell, Botkin, Fry, Wright, & Jarvis, 1968; Selman, 1971). Early tests using global cognitive constructs to predict social behavior produced mixed findings (Shantz, 1975, 1983). In the 1970s, theories of information processing were introduced by researchers such as Newell and Simon (1972). The new approaches focused on specific components or steps of *on-line* (or real-time) cognition rather than global cognitive constructs. This perspective of *real-time cognition* quickly gained popularity, and led to major changes in empirical and theoretical approaches to the study of social cognition in children, of which Crick and Dodge were among the major contributors (Crick & Dodge, 1994; Dodge, 1986). By specifying the information-processing steps in which children engage when faced with social situational cues, Crick and Dodge's (1994) SIP model constituted a substantial advancement in the understanding of children's social adjustment. Arguably, Crick and Dodge's SIP model has become the major theoretical and primary empirical approach to the study of how social cognition affects child behavior (Arsenio & Lemerise, 2010).

The SIP model was formulated as a global framework representing cognitive operations underlying social behavior, and the primary application of the SIP model has been toward understanding aggressive behavior in children. Many empirical studies have demonstrated the relationship between processing patterns at each SIP step and aggressive behaviors. Specifically, research has shown that as compared with their nonaggressive peers, aggressive children encode fewer and less-benign social cues (Step 1; Dodge & Newman, 1981; Gouze, 1987); attribute more hostile intentions to other's actions (Step 2; Feldman & Dodge, 1989); select goals that damage relationships (Step 3); generate fewer and less prosocial responses (Step 4; Pettit, Dodge, & Brown, 1988); and evaluate aggressive responses more favorably and expect more positive outcomes from aggressive behavior (Step 5).

Drawing primarily from the implications of SIP theory and the empirical findings regarding the association between SIP patterns and children's aggressive behaviors, the *Making Choices* curriculum was developed to strengthen skills in processing social information and regulating emotions. The curriculum consists of seven modules or units of lessons. The first unit is devoted to understanding and regulating emotions, and the last unit is about enacting a selected strategy. The other five units correspond to the five SIP steps in sequence (i.e., encoding social and environmental cues, interpreting cues and intentions, setting relational goals, formulating alternative social strategies, selecting prosocial strategies; Fraser, Day, Galinsky, Hodges, & Smokowski, 2004; Fraser et al., 2005).

Evaluation of *CSP*

The evaluation study of *CSP* took place over a 3-year study period (2004 to 2007). The study used a sampling strategy, consent process, core measures, and random assignment procedure that were common to the seven SACD programs included in the national evaluation. The evaluation team initially recruited 10 schools in two rural North Carolina school districts in 2004 (the 2004 cohort). The 10 schools were matched into pairs based on five school-level characteristics: school size, third-grade class size, ethnic composition, math and reading achievement scores, and rate of participation in the federal free and reduced priced lunch program. Within each pair, schools were randomly assigned to either the treatment or the control condition. In the second year of the study, four schools were added to increase the power of analysis in 2005 (the 2005 cohorts); these additional schools were matched and assigned to treatment conditions using the same procedures described above.

The treatment condition included delivery of the *Making Choices* social skills curriculum to all grades (i.e., Grades 1 through 5) throughout the 3-year project period. However, the efficacy test focused on the 2004 and 2005 cohorts of third-grade students. Students in the treatment condition received 28 *Making Choices* core lessons in their third-grade year, and eight *Making Choices* follow-up or "booster shot" lessons in each of the fourth- and fifth-grade years. At the outset of each school year, teachers and support staff were trained to deliver the *Making Choices* curriculum and to use strategies for classroom management that focused on managing peer social dynamics and behavior problems. In addition, teachers were provided with consultation and training support through biweekly grade-level meetings held throughout the school year.

Outcomes were assessed in the fall and spring of each year during the 3-year study period. *CPS* also collected data on program fidelity, including the number of lessons and minutes of *Making Choices* curriculum taught in each classroom. The current dosage analysis used the length of treatment exposure (i.e., the number of minutes) as a measure of dosage (see Method section for discussion of rationale for this measure).

Study Hypotheses

Combining the core social-skills training curriculum with classroom behavior-management strategies, *CSP* was expected to promote social competence and reduce aggressive behavior through enhancing children's emotional regulation skills and social information processing skills. Findings from a previous analysis of data from the 2004 cohort suggested that the intervention produced negative effects in the third grade and positive cumulative effects in the fourth and fifth grades on multiple dimensions of outcomes including social competence, classroom behavior, and academic achievement (Fraser et al., 2009).

The current analysis advances efforts to evaluate *CSP* by testing whether the intervention effects varied by the dosage (i.e., length of treatment exposure). Preliminary analysis has shown that the implementation of the *Making Choices* curriculum varied greatly across classrooms, ranging from 268 to 2,340 minutes in the third grade and from 220 to 600 minutes in the fourth grade. Because the content of the *Making Choices* curriculum is sequenced corresponding to SIP steps, it is important that students receive all the lessons. Therefore, it is reasonable to expect that greater gains would be observed in classrooms in which the program was fully implemented. However, this prediction would be valid only if the same quality is associated with each of the training minutes, which was not the case in the *CSP* study. Based on follow-up information, the meaning of a *minute* of program exposure is not equivalent across classrooms. For example, one classroom reported 2,340 minutes of instruction, which was nearly 2 times the length of training prescribed by the intervention design (i.e., 1,120 minutes). Follow-up information revealed that the classroom teacher reported minutes of using *Making Choices* stories for her other subject areas (e.g., language arts) with the same classroom of students. Moreover, the same program content might be delivered with varying quality, resulting in differential responsiveness of participants. Quality of delivery and participant responsiveness are crucial factors affecting program outcomes (Dane & Schneider, 1998). Although the study did not collect data on these measures, follow-up teacher interviews suggest that the *Making Choices* content was not delivered with the same quality across classrooms. Taking into account both the quantity and the quality of implementation, it is hypothesized that the observed relationship between the number of minutes and outcomes would be nonlinear, although a general trend could be observed.

Method

Analysis Sample: Inclusion Criteria

This analysis used data from both the 2004 cohort and the 2005 cohort of third-grade students. For the purpose of this dosage analysis, only participants assigned to the treatment condition were included. The exclusion of the control group is due to both substantive and statistical considerations. First, the primary goal of a dosage analysis is to identify optimal doses and not to evaluate the overall effects of a treatment (for overall effects of *Making Choices*, see Fraser et al., 2009). Second, this dosage analysis treats the treatment variable as continuous and assumes a normal distribution of the observations (Hirano & Imbens, 2004); therefore, including control participants with zero minutes of treatment would violate the normality assumption.

One treatment school in the 2005 cohort withdrew from the study a year after the *CSP* program was implemented, and was excluded from this study. During the 3-year study period, some students changed schools across the treatment conditions, and other students entered or left participating schools. Therefore, the following three criteria were established for including a student in the analysis: (a) students who moved from a comparison school to a treatment school were treated as *enterers,* whereas those who moved from a treatment school to a comparison school were excluded; (b) students who left the study were retained in the analysis until their attrition date; (c) students who entered the study after third grade were excluded because these students would not have received the core 28 *Making Choices* lessons. This analysis defined enterers by date the student was first on the class roster (i.e., the time they were first enrolled in a classroom), rather than the date they were consented to the study because these students would have received the classroom lessons from their date of first entry into the class. The final sample consisted of 400 students from 30 classrooms (173 Black, 155 White, 30 Hispanic, 14 American Indian, and 28 other ethnicity), including 323 students in 23 classrooms from the 2004 cohort, and 77 students in seven classrooms from the 2005 cohort. The majority of the sample was female (55%, n = 220), and 45% was male (n = 180).

Missing Data

As is common in longitudinal studies, missing data occurred at each wave of data collection, ranging from 14% to 18% on the outcome variables. Cases with missing data were deleted. The use of listwise deletion other than any type of imputation method was chosen based on the focus of the current analysis. Dosage analysis is essentially a type of treatment-of-the-treated (TOT) analysis, in which the effects of interventions are estimated based on differential program exposure and implementation. The primary concern in this type of study is internal validity, and not external validity or the generalizability of the findings.

Outcome Measures

The current dosage analysis focused on a set of key outcomes, including social competence, emotional regulation, relational aggression, overt aggression, and four SIP skill measures. The selection of the key outcomes was consistent with the primary programmatic goal of the *CSP* program: to promote social competence and reduce aggressive behavior by enhancing emotional regulation skills and altering SIP patterns that lead to aggressive behavior (Fraser et al., 2009).

Social competence. Social competence can be broadly conceived as the capacity to integrate cognition, affect, and behavior to achieve specified social tasks and positive developmental outcomes (Waters & Sroufe, 1983; Weissberg & Greenberg, 1998). *CSP* measured social competence using the Carolina Child Checklist-Teacher Form (CCC-T; Macgowan, Nash, & Fraser, 2002). The CCC-T uses a 6-point Likert-type scale with response options ranging from *never* (0) to *always* (5). The CCC-T is intended for observations of children between ages 6 to 12 years, and measures five dimensions of behavior including cognitive concentration (e.g., concentrates in class), social contact (e.g., plays with others), authority acceptance (e.g., breaks rules), and social competence. Consistent with the content of *Making Choices*, the CCC-T has two subscales: emotional regulation and prosocial behavior. Therefore, the social competence measure comprises items related to emotional regulation (e.g., controls temper when there is a disagreement, can calm down when excited or all wound up) and prosocial behavior (e.g., resolves peer problems on his or her own; Macgowan, Nash, & Fraser, 2002). The Cronbach's alpha for this measure was .93.

Emotional regulation. Emotional regulation refers to a person's ability to manage his or her emotions. Children's ability to regulate their emotions was measured using the emotion subscale of the CCC-T. This subscale assesses emotion management using items such as "can calm down when excited or all wound up"' and "controls temper when there is a disagreement." The emotional regulation subscale has a Cronbach's alpha of .85**.**

Social aggression. In the evaluation of *CSP***,** *social aggression* refers to manipulating group acceptance by excluding or attacking the character of another person (Cairns, Cairns, Neckerman, Ferguson, & Gariepy, 1989). In the literature, social aggression is often used interchangeably with terms such as *relational aggression* and *indirect aggression* (e.g., Card, Stucky, Sawalani, & Little, 2008). However, social aggression has "wider features of social coercion and verbal confrontation that may not be intended solely to harm relationships with peers but to connote efforts to control and establish status" (Fraser et al., 2005, p. 1048).

Social aggression was measured using a subscale of the CCC-T, which was labeled the relational aggression subscale (e.g., Fraser et al., 2004; Macgowan et al., 2002; Smokowski et al., 2004) for consistency with the scale from which the subscale was adapted (i.e., the Relational Victimization subscale derived from the Social Experience Questionnaire; Crick & Grotpeter, 1996). Following the approach of Fraser et al. (2005), this article uses the more inclusive term of social aggression to refer to this subscale.

The social aggression subscale consists of nine items such as "excludes other kids from peer group," "lies to make peers dislike a student," and "stubborn." These items capture a range of behavioral characteristics consistent with relational victimization, verbal aggression, bullying, and authority avoidance (Fraser et al., 2005). The item that measures authority avoidance was included because stubbornness characterizes an early-start delinquency trajectory (Thornberry et al., 2004; Underwood, 2003). The social aggression subscale has a Cronbach's alpha of .91.

Overt aggression. Overt aggression involves the use of direct, confrontational behaviors that are intended to harm another through physical damage or the threat of such damage (e.g., shoving, kicking, threatening to beat up a peer; Grotpeter & Crick, 1996). In the *CSP* study, overt aggression was measured using the aggression subscale of the Interpersonal Competence Scale—Teacher (ICST). The ICST is a 21-item, teacher report questionnaire that assesses social and behavioral characteristics of children on a 7-point Likert scale from *never* to *always* (Cairns et al., 1995). The aggression subscale consists of items that indicate overt physical and verbal aggression (e.g., gets into trouble, gets into fights, argues). Cronbach's alpha for the aggression subscale is .82.

SIP skills. SIP skills were measured using the Skill Level Activity (SLA), which is an adaptation of Dodge's (1980) Home Interview that assesses attributional bias; the instrument is designed for group administration as a pen-and-paper measure (for more information on SLA, see Day, 2004; Fraser et al., 2005). The SLA uses a story-based child assessment protocol. Each of the six short stories describes a situation in which a peer interaction of ambiguous intent occurs. Following each of the stories, students are asked to respond to four questions that correspond to encoding cues ($\alpha = .78$), attributing (hostile) intent ($\alpha = .52$), formulating prosocial goals ($\alpha = .76$), and making a response decision ($\alpha = .80$).

Analytic Strategy

The data analyses involved several critical decisions regarding the dosage measure, assessment points, and statistical methods. The decisions made and the rationale for making each of those decisions are discussed below. A brief introduction to dosage statistical methods is also provided.

Rationale for choosing number of minutes as the dosage measurement. To measure the program dosage of an intervention such as the *Making Choices* curriculum, the researcher can choose either the number of lessons taught or cumulative minutes of lessons delivered. Each of the *Making Choices* lessons was designed to be delivered in one 40-minute session. Therefore, the 28 third-grade lessons and 8 fourth-grade lessons would be delivered in 1,120 and 320 minutes, respectively. However, in practice, the total minutes of *Making Choices* instruction delivered in the third-grade year ranged from 268 to 2,340 minutes across classrooms (M = 1,071.73, Mdn = 1,088, SD = 385.67). The average minutes per lesson in Grade 3 classrooms varied from 32.7 minutes to 83.6 minutes. Similar variation was observed in the Grade 4 classrooms, with the total minutes ranging from 220 to 600 (M = 360, Mdn = 374.25, SD = 77.14) and the average minutes per lesson ranging from 31.4 to 75.0 minutes. Although it is important for students to receive all the lessons in order to master the whole set of problem-solving skills, the actual time investment in each lesson is also an important factor. Therefore, cumulative minutes by grade were used as the measure of dosage.

Assessment points. Data from both the 2004 and 2005 cohorts were used. The 2005 cohort did not have data at Grade 5 because it joined the study a year after the study started, and the study ended before this cohort entered Grade 5 (Fraser et al., 2009). Therefore, the dosage effects could be estimated only for third graders and fourth graders when data from the 2004 and 2005 cohorts were combined. The analyses in the present study estimated the dosage effects at the end of Grade 3 and Grade 4, which represented educationally relevant traditional assessment points. Because students changed classrooms as they moved from Grade 3 to Grade 4, the dosage effects were not estimated as a function of the cumulative minutes over 2 years (i.e., Grade 3 minutes + Grade 4 minutes). Using cumulative minutes over Grade 3 and Grade 4 would result in a very small number of students at each minute level. When the number of observations associated with each minute level is too small, the average dosage effects estimated at each dosage level (i.e., minute level) would not be meaningful.

Change score versus point score. To control for rater effects, change scores of the outcomes within Grade 3 and Grade 4 were used as the dependent variable. Different raters might give different scores on the same behavior; some raters are liberal, some raters are strict. Differences in point scores might reflect differences between raters, and therefore, might not be a faithful measure of the actual difference in behavior (Guo & Hussey, 1999). Assuming that raters rate consistently over time, using change scores can remove the rater effects.

Rationale for selection of GPS method. The decision to use a GPS method rather than a conventional linear regression model was based on the advantages offered by the GPS method. First, a GPS summarizes information in confounding covariates. By comparing the distribution of the GPS, it is immediately obvious whether the groups under comparison have overlapping distributions of observed covariates. The overlap of GPS (i.e., the common support region) indicates the range over which the data will support estimates of the treatment effects, and the analysis is restricted to participants whose GPS fall in the common support region (for more information about common support region, see Li, 2012). In contrast, assessing the areas of common support among dosage groups is not feasible when using the conventional linear regression method.

Second, similar to other methods in the family of propensity score modeling, GPS modeling is robust to model misspecification (Drake, 1993) because propensity scores serve the purpose of balancing groups. As long as balance is achieved, incorrect modeling is not an issue of concern (Williamson, Morley, Lucas, & Carpenter, 2012). In contrast, linear regression models depend on the specific form of the model to extrapolate estimates of treatment effects. Results from outcome regression models are sensitive to model specification, which is particularly a problem when the covariate distributions in dosage groups are very different (Drake, 1993; Intosh & Rubin 1999; Perkins, Tu, Underhill, Zhou, & Murray, 2000; Rubin, 1997).

Third, the balance achieved through applying GPS allows a conditional causal interpretation of the findings. Without means to address the issue of initial group equivalence, regression analyses are essentially correlational (Guo & Fraser, 2010).

Last, GPS modeling is advantageous for practical reasons. A single set of GPS can be used for evaluation of more than one outcome. This feature can be important when there are many outcomes of interest. Regression modeling requires fitting individual regression models for each outcome, which is time-consuming (Zanutto, Lu, & Hornik, 2005).

GPS-based method with continuous treatment. The GPS method with continuous treatment is a relatively recent development in the family of propensity score-based methods. Propensity score-based methods have evolved from the initial methods with two treatment levels (Rosenbaum & Rubin, 1983) to multiple treatment levels (Imbens, 2000; Joffe &

Rosenbaum, 1999) to the recent development of the GPS method with continuous treatment (Hirano & Imbens, 2004).

In this analysis, the use of the GPS method with continuous treatment consisted of five steps. In the first step, the conditional distribution of the treatment (*T;* i.e., minutes) given covariates $\boldsymbol{X_i}$ was estimated. It is assumed that the treatment or its transformation has a normal distribution conditional on the covariates:

$$g(T_i) \mid X_i \sim N\{(\beta_0 + \beta_1^{'} X_i), \sigma^2\},$$

where $g(T_i)$ is a transformation of the treatment variable (i.e., minutes) that can satisfy the normality assumption. In this analysis, a log transformation of minutes is applied to satisfy the normal distribution assumption. Parameters β_0, β_1, and σ^2 are estimated using maximum likelihood.

The selection of covariates $\boldsymbol{X_i}$ is a key issue in applying any GPS methods. In this analysis, the initial selection of covariates was based on the theoretical and empirical association of each variable with the treatment (i.e., minutes of instruction) and the outcomes (for a review on criteria for the selection of covariates see Li, 2012). The final decision on the inclusion of covariates and their higher order terms was made through iterations of specifying the model that estimates the conditional distribution of minutes, estimating GPS, testing covariate balance, and respecifying models that estimate the conditional distribution of minutes.

In the second step, the GPS was estimated by modeling the conditional density of the log transformation of minutes given covariates using a simple normal density function:

$$\hat{R}_i = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} [g(T_i) - (\hat{\beta}_0 + \hat{\beta}_1^{'} X_i)]^2 \right)$$

where $\hat{\beta}_0$, $\hat{\beta}_1^{'}$, and $\hat{\sigma}^2$ are parameters estimated from the first step; $(\hat{\beta}_0 + \hat{\beta}_1^{'} X_i)$ is the conditional expectation of treatment (i.e., the predicted value of treatment).

In the third step, the common support region was identified and balance was tested by applying the estimated GPS. To identify the common support region, I used the approach outlined by Flores et al. (2010). The sample was divided into three dosage subgroups of approximately equal size. The cut points were 1,030 minutes and 1,105 minutes. Three sets of GPS were then estimated at the median of each dosage interval. The common support region with respect to each interval was obtained by comparing the GPS of participants belonging to the interval and those not belonging to the interval. The analysis sample was then limited to participants whose GPS simultaneously occurred in the three common support regions.

To test covariate balance, I utilized a common approach by regressing each covariate on the treatment variable. A linear regression model was used to test the balance of continuous covariates (e.g., adequate yearly progress, social competence). For binary covariates (e.g., gender, ethnicity), I used a logistic regression model. Three sets of tests were conducted. The first test was done with the full sample before applying common support. After applying common support (i.e., limiting the sample to children whose GPS fell in the common support region), the second and third sets of tests were conducted without and with conditioning on the estimated GPS.

In the fourth step, the conditional expectation of the outcome was estimated. In Hirano and Imbens’ (2004) approach, the conditional expectation of the outcome was estimated as a flexible function of two scalar variables (i.e., the treatment variable and the estimated GPS). This approach is suitable in a situation in which all the covariates are balanced after applying the GPS; however, this was not the case in the current analysis. After applying the GPS, some covariates remained unbalanced. To address the residual unbalance in covariates, I followed the approach used in Abadie and Imbens (2002) and Lechner and Melly (2010) and I included the unbalanced covariates in the regression model that estimates the conditional expectation of the outcome (i.e., the dose-response model). The model can be written as:

$$E[Y_i \mid T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 \hat{R}_i + \alpha_3 T_i \hat{R}_i + \alpha_4 X_{1i} + \ldots + \alpha_n X_{ni}.$$

The parameters were estimated by ordinary least squares.

In the fifth step, the average potential outcome for each minute level of interest (*t*) was estimated. The estimation was done by averaging the conditional expectation over the estimated GPS at the particular minute level of interest using the coefficients estimated in Step 4:

$$E[Y(t)] = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha}_0 + \hat{\alpha}_1 \cdot t + \hat{\alpha}_2 \cdot \hat{r}(t, X_i) + \hat{\alpha}_3 \cdot t \cdot \hat{r}(t, X_i) + \hat{\alpha}_4 X_{1i} + \hat{\alpha}_n X_{ni}\right),$$

where $\hat{r}(t, X_i)$ is the estimated GPS at treatment level *t* given X_i. I chose 10 treatment levels, which included the lowest treatment level and the treatment levels that included approximately 10% to 100% of participants.

Results

Covariates Included in the GPS Model.

Through iteration of specifying the GPS model, checking balance, and respecifying the GPS model, an optimal GPS model that included 26 linear terms and nine square terms was identified. These covariates included variables measured at student, classroom, and school levels.

Estimated GPS and Common Support Region

Shown in Table 3.1a and Table 3.1b, the overall estimated GPS for Grade 3 (i.e., the conditional density of the treatment given the covariates) ranged from 0.0004511 to 3.677995, with a mean of 2.722791 and a standard deviation of 0.97603; the overall estimated GPS for Grade 4 ranged from 0.0128262 to 2.834811, with a mean of 2.123233 and a standard deviation of 0.8406825. To identify common support regions and check balance, three sets of GPS (i.e., GPS_1, GPS_2, and GPS_3) were estimated at the median of each of the three dosage intervals. For Grade 3, the three treatment intervals were minute <1030, $1030 \leq$ minute <1123, and minute ≥ 1123. For Grade 4, the three treatment intervals were minute$\leq$330, 330$<$minute$\leq$385, and minute$>$385. A common support region with respect to each dosage interval was obtained by comparing the GPS of participants belonging to the interval and those not belonging to the interval. The three common support regions defined by each set of the GPS are shown in Table 3.2a for Grade 3 and Table 3.2b for Grade 4.

Table 3.1a

Overall GPS and the GPS Estimated at the Median of Each Treatment Interval –Grade 3

GPS		*N*	*M*	*SD*	Min.	Max.
Overall GPS		267	2.722791	.97603	.0004511	3.677995
GPS estimated at median of each treatment interval	GPS_1	267	.8512852	.933295	2.70e-29	3.607073
	GPS_2	267	2.159365	1.442738	2.10e-38	3.678463
	GPS_3	267	1.648196	1.648196	0	3.678609

Table 3.1b

Overall GPS and the GPS Estimated at the Median of Each Treatment Interval – Grade 4

GPS		*N*	*M*	*SD*	Min.	Max.
Overall GPS		253	2.123233	.8406825	.0128262	2.834811
GPS estimated at median of each treatment interval	GPS_1	253	1.802199	.9759995	.0476953	2.834809
	GPS_2	253	1.896272	.634314	.2109219	2.834792
	GPS_3	253	1.374272	1.073734	.0109267	2.834721

Table 3.2a

Common Support Region—Grade 3

Treatment interval with GPS estimate	Dosage group	GPS_1	
		Minimum	Maximum
≤1030 (Median = 966)	Minute≤1030	2.70e-29	3.607073
	Minute >1030	.0002477	3.584874
Common support region 1: [.0002477, 3.584874]			
		GPS_2	
		Minimum	Maximum
1031-1105 (Median = 1076)	1030 <Minute≤1105	.0166222	3.678462
	Minute<1030 & Minute >1105	2.10e-38	3.678357
Common support region 2: [.0166222, 3.678357]			
		GPS_3	
		Minimum	Maximum
>1105 (Median = 1234)	Minute >1105	.0004511	3.678609
	Minute ≤1105	0	3.676587
Common support region 3: [.0004511, 3.676587]			

Table 3.2b

Common Support Region—Grade 4

Treatment interval with GPS estimate	Dosage group	GPS_1	
		Minimum	Maximum
≤330 (Median = 330)	Minute ≤330	.3844731	2.834809
	Minute >330	.0476953	2.833984
Common support region 1: [.3844731, 2.833984]			
		GPS_2	
		Minimum	Maximum
331-385 (Median = 375)	330<Minute≤385	.7244377	2.829402
	Minute≤330 & Minute >385	.2109219	2.834792
Common support region 2: [.7244377, 2.829402]			
		GPS_3	
		Minimum	Maximum
>385 (Median = 435)	Minute >385	.8048868	2.834721
	Minute ≤385	.0109267	2.834246
Common support region 3: [.8048868, 2.834246]			

The analysis sample was then limited to participants whose GPS simultaneously occurred in the three common support regions. This reduced the sample size to 216 participants for Grade 3 and 196 participants for Grade 4. Participants with extremely low or high minute reports fell outside of the common support region (i.e., third graders with minute reports of 268, 285, 351, 585, 1640, and 2340 minutes; fourth graders with minute reports of

220, 285, and 600 minutes). To explore the difference between the full sample and the analytic sample after applying the common support region, the mean change score of outcomes were computed for the full sample and the final analytic sample. As shown in Table 3.3, the full sample and the final analytic sample were similar on the average change score of most of the outcome variables. The absolute value of difference between the mean score of the full sample and the mean score of the analysis sample exceeded .1 only for one variable (i.e., Grade 3 hostile attribution).

Table 3.3
Unadjusted Means (Standard Deviations) of Outcome Variables

Outcome variable	*M* (*SD*)					
	M change score within Grade 3			*M* change score within Grade 4		
	Full sample (f1)	Analysis sample (a1)	Difference (f1-a1)	Full sample (f2)	Analysis sample (a2)	Difference (f2-a2)
Social competence	-.021 (.728)	.026 (.729)	-.047	.044 (.725)	.130 (.597)	-.086
Emotional regulation	.003 (.786)	.038 (.792)	-.035	.037 (.802)	.130 (.651)	.093
Social aggression	-.221 (.653)	-.179 (.629)	-.042	-.048 (.639)	-.032 (.635)	-.016
Overt aggression	.206 (1.13)	.197 (1.13)	.009	.001 (.990)	.087 (1.066)	-.086
Cue identification	-.040 (.211)	-.050 (.214)	.010	.027 (.247)	.012 (.246)	.015
Hostile attribution	-.101 (.336)	.001 (.334)	-.102	.017 (.352)	.017 (.356)	0
Goal formulation	.017 (.309)	.017 (.300)	0	-.004 (.401)	.060 (.377)	-.064
Response decision	.021 (.359)	.012 (.344)	.009	-.061 (.425)	-.023 (.402)	-.038

Balance Check

Results from the three sets of balance tests for Grade 3 and Grade 4 are shown in Table 3.4a and Table 3.4b. The table includes only the variables that were unbalanced (i.e., $p < 0.1$). Using data from the third graders, 10 covariates were unbalanced before applying common support. The number of unbalanced covariates was reduced to four after applying common support (i.e., limiting the sample to children whose GPS fell in the common support region) but without accounting for the estimated GPS. There was no further reduction on the number of unbalanced covariates by accounting for the estimated GPS. It is worth noting that one covariate (i.e., hostile attribution) that was balanced initially became unbalanced after applying common support. Using data from the fourth graders, seven covariates were unbalanced before applying common support. After applying common support, the number of unbalanced covariates was reduced to two. When accounting for the estimated GPS, the number of unbalanced covariates remained unchanged. However, the two covariates were not the same before and after accounting for the estimated GPS—one initially balanced covariate became unbalanced (i.e., number of weeks devoted to tolerance and diversity activities), and one initially unbalanced covariate became balanced (i.e., percentage of students receiving free or reduced lunch).

Table 3.4a

Results of Balance Check-Grade 3

Covariate	Before Applying Common Support		After Applying Common Support			
			Without GPS		With GPS	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Demographic						
Hispanic	-.0019007	.001**	-.0003067	.929	-.0004128	.906
Student Report						
Cognitive conc.	.0006413	.004**	.0010989	.062^{+}	.0011028	.062^{+}
Academic achievement	.0007143	.025*	-.000438	.592	-.0003505	.669
Encoding	.0001771	<.001**	.0000328	.760	.0000307	.776
Hostile attribution	.0000315	.612	.0004526	.007**	.0004628	.006**
Teacher Report						
WOTD [a] *(s0_b4_wk?)*	.0012526	<.001**	.0080302	<.001**	.0079101	.<001**
PSS [b]	.0004057	.011*	.0004651	.335	.0004264	.378
Professional interest	.000818	<.001**	.0005689	.017*	.0005853	.014*
School Report						
PFRL [c]	-.0000853	.004**	.00002	.828	.0000238	.798
Adequate yearly progress	.0113292	<.001**	-.0004432	.847	-.0006342	.783
Income to poverty ratio	.078094	<.001**	.0922802	.0157	.0914177	.163

Note. a WOTD = Weeks devoted to tolerance and diversity activities. *b* PSS = Perception of student support. *c* PFRL = Percentage of students receiving free or reduced lunch. ** $p < .01$, * $p < .05$, $^{+}$ $p < .10$, two-tailed

Table 3.4b

Results of Balance Check-Grade 4

Covariate	Before Applying Common Support		After Applying Common Support			
			Without GPS		With GPS	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Demographic						
Hispanic	.0051156	.037*	-.0066041	.456	-.0062936	.519
Student Report						
Academic achievement	-.0020823	.078^{+}	.00148	.529	.0006423	.818
Teacher Report						
WOTD [a]	.00406030	<.001**	.0040929	.184	.0071311	.050^{+}
WORP [b]	-.0004012	.054^{+}	-.0003785	.220	-.0005867	.109
WOBM [c]	-.006474	.017*	-.0070878	.236	-.0041777	.555
PSS [d]	.0023919	<.001**	.0011887	.277	.0011821	.363
School Report						
PFRL [e]	-.0005676	<.001**	-.0004157	.047*	-.0004051	.103
Adequate yearly progress	.0111947	.009**	.0235103	.046*	.0278629	.046*

Note. a WOTD = Number of weeks devoted to tolerance and diversity activities. *b* WORP = Number of weeks devoted to risk behavior prevention. *c* WOBM = Number of weeks devoted to behavior management programs. *d* PSS = Perception of student support. *e* PFRL = Percentage of students receiving free or reduced lunch. ** $p < .01$, * $p < .05$, $^{+}$ $p < .10$, two-tailed

Conditional Expectation of the Outcome

Tables 3.5a and 3.5b present the coefficients and associated *p*-values resulting from modeling the conditional expectation of the outcomes (i.e., change scores within Grade 3 and Grade 4) as a function of the treatment variable (i.e., minutes), the estimated GPS, and the unbalanced covariates. In all, eight models were estimated for each grade, each corresponding to one of the eight outcome variables. As Hirano and Imbens (2004) emphasized, the estimated coefficients were not directly interpretable. However, the *p*-values are meaningful; the *p*-values associated with the treatment can be interpreted as indicating whether the treatment had significant effects on the outcomes. Moreover, the *p*-values associated with the GPS and the unbalanced covariates indicate whether the covariates summarized in the GPS and the unbalanced covariates introduced bias.

Table 3.5a

Dose Response Function of Outcomes-Grade 3

Independent Variable	Social Competence (n=210)		Emotional Regulation (n = 210)		Social Aggression (n = 210)		Overt Aggression (n = 210)		Cue Identification (n = 201)		Hostile Attribution (n = 201)		Goal Formulation (n = 202)		Response Decision (n = 202)	
	Coef.[a]	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value
Minute	-.005	.001**	-.004	.040*	.000	.003**	.005	0.063^{+}	.000	.524	.000	.978	-.001	.061^{+}	-.001	.541
GPS	-1.979	.003**	-1.425	.051^{+}	.132	.002**	1.788	.110	.132	.542	.040	.775	-.789	.007**	-.317	.353
Minute-GPS	.0020	.001**	.0015	.021*	-.000	.001**	-.002	.106	-.000	.630	-.000	.619	.001	.009**	.000	.493
Cognitive Concentration	-.135	.003**	-.147	.003**	-.005	.757	.138	.067^{+}	-.005	.739	-.034	.046*	.028	.154	.003	.907
Hostile Attribution	-.258	.101	-.329	.058^{+}	-.079	.304	.093	.727	-.079	.125	-.667	<.001**	.093	.181	.092	.256
WOTD[b]	-.037	.338	-.063	.141	-.009	.030	-.012	.854	-.0092	.461	.014	.284	-.061	<.001**	-.030	.127
Professional Interest	.363	.002**	.377	.003**	.049	.004	-.224	.252	.049	.190	.032	.476	.035	.493	-.014	.808
Constant	4.178	0.023*	2.413	.229	-.621	.055	-4.781	.121	-.621	.303	.383	.232	1.268	.118	.815	.391

Note. [a] Coef = Coefficient. [b] WOTD=Number of weeks devoted to tolerance and diversity activities. ** p < .01, * p < .05, $^{+}$ p < .10, two-tailed.

Table 3.5b *Dose Response Function of Outcomes-Grade 4*

Independent Variable	Social Competence (n=186)		Emotional Regulation (n = 186)		Social Aggression (n = 186)		Overt Aggression (n = 186)		Cue Identification (n = 175)		Hostile Attribution (n = 175)		Goal Formulation (n = 176)		Response Decision (n = 176)	
	Coef.[a]	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value	Coef.	*P* value
Minute	.001	.490	.002	.242	.000	.645	-.002	.474	-.001	.129	.001	.397	-.001	.590	-.001	.496
GPS	.073	.883	.657	.237	.408	.443	-.498	.577	-.689	.001**	.451	.169	.036	.916	-.041	.912
Minute-GPS	.000	.947	-.001	.353	-.001	.432	.001	.614	.002	.001**	-.001	.183	-.000	.836	-.000	.972
PFRL[b]	-2.556	.001*	-2.245	.009**	-1.463	.073^{+}	3.655	.008**	.457	.158	.672	.185	-.790	.140	-.366	.525
Adequate yearly progress	-.031	.027*	-.026	.099^{+}	.006	.677	.020	.427	-.003	.687	.010	.339	-.004	.695	-.001	.956
WOTD[c]	-.040	.194	-.020	.556	-.039	.241	.062	.267	-.006	.611	.007	.734	-.000	.998	-.015	.499
Constant	3.605	.042*	2.360	.227	-.032	.986	-2.747	.382	.492	.533	-1.695	.173	1.155	.376	.685	.627

Note. [a] Coef = Coefficient. [b] PFRL = Percentage of students receiving free or reduced lunch. [c] WOTD = Number of weeks devoted to tolerance and diversity activities. ** p < .01, * p < .05, $^{+}$ p < .10, two-tailed

According to the *p*-value associated with minutes, the intervention had significant dosage effects at the end of Grade 3 on social competence (p = .001), emotional regulation (p =.040), and social aggression (p = .003) at the alpha level of .05. Using the criterion of alpha level of .10, dosage effects were also significant on overt aggression (p = .063) and goal formulation (p = .061). No significant dosage effects were found at the end of Grade 4. For outcomes on which the intervention had significant overall dosage effects, an additional analytical step was taken to compute the average dosage effect at each level of the treatment of interest.

Average Dosage Effect at Each Level of the *Making Choices* Program

With the coefficients estimated by modeling the conditional expectation of the outcomes, average effects by dosage levels were estimated for significant outcomes by averaging the estimates of participants with the same dosage (i.e., number of minutes). The current study estimated the average dosage effects at 10 treatment levels, which included the lowest treatment level and the treatment levels that included approximately 10% to 100% of participants. Results are reported in Table 3.6. The average dosage effects of social competence and emotional regulation at the lowest minute level (minute = 906) were negative (ES = -.6316 and -.6664). However, positive effects emerged as the minutes of treatment exposure increased. Before the minutes reaches a level that far exceeds the designed level of 1120 minutes (e.g., minute = 1,380), a trend relationship exists of average dosage effects and the number of minutes; however, the relationship is not strictly linear. The average dosage effects at the highest level (i.e., minute = 1,380) were negative. The results do not show a trend relationship of effects on social aggression and overt aggression with levels of minutes. It is noteworthy that the average dosage effects at most of the minute levels are positive for social aggression, but negative for overt aggression. As suggested by the large *p*-value associated with minutes for the outcome of goal formulation, the average dosage effects are close to zero at most of the dosage levels.

Table 3.6
Average Dosage Effects – Grade 3

Intervention Dose (minutes of instruction)	Average Outcome at Each Treatment Dose				
	Social Competence (+)[a]	Emotional Regulation (+)	Social Aggression (-)	Overt Aggression (-)	Goal Formulation (+)
906	-.6316	-.6664	-.6683	.5617	-.1119
945	.0658	.0036	-.1379	-.0126	-.0016
1030	.0502	.0349	-.1278	.0891	.0255
1068	.1332	.1571	-.0928	.1183	.0281
1088	.0652	.1191	-.1513	.2683	.0045
1105	.1363	.1823	-.0976	.1791	.0273
1151	.2624	.3182	.0287	.1728	.0731
1234	.1372	.2045	.0431	.2166	.1405
1292	.3851	.3100	-.1374	-.1126	-.1735
1380	-.2943	-.1920	-.3079	.5593	.1683

Note. *a* = hypothetical sign

Discussion

Findings from this dosage analysis suggest that the intervention had overall dosage effects at the end of Grade 3 on social competence, emotional regulation, and social aggression at alpha level of 0.05. Dosage effects were also significant on overt aggression and goal formulation at alpha level of .10. No dosage effects were observed in Grade 4. A clear pattern of dosage effects in Grade 3 was observed only on social competence and emotional regulation, although the relationship between increases in social competence and emotional regulation and the increases in treatment exposure (i.e., minutes) was not linear.

Exploratory in nature, the findings are both promising and puzzling. Although a dosage effect was observed, the effects varied across levels (e.g., from –.6316 to .3851 for social competence). The effects also varied for participants with similar dosages. For example, the difference between 1234 and 1292 is 58 minutes, which might be reasonably viewed as negligible. However, the difference of average dosage effects was .2479 (i.e., .3851 - .1372), which is substantial. These findings suggest an important area that needs further examination—the quality of implementation.

Quality of implementation is part of the fidelity or integrity issues that have been discussed widely by researchers, although with varying degrees of depth and intensity (e.g., CPPRG, 1999; Dane & Schneider, 1998; Fraser et al. 2009; Fraser et al., 2011; Schoenwald et al., 2011). Fidelity or integrity refers to the extent to which specified procedures and activities are implemented as planned (e.g., Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Moncher & Prinz, 1991). Discussed previously, implementation has quantitative as well as qualitative aspects (Scheoenwald et al., 2010; Li, Fraser, & Wike, 2012). High fidelity is widely recognized as leading to better outcomes (e.g., Battistich, Schaps, Watson, & Solomon, 1996; Botvin, Baker, Dusenbury, Tortu, & Botvin, 1990; CPPRG, 1999; Pentz et al., 1990). However, more attention has been given to the quantitative aspects of fidelity than to the qualitative aspects of fidelity (Guo & Fraser, 2010; Li et al., 2010). When examining only the quantitative aspects of fidelity (e.g., the length of program exposure), similar to the current study, researchers sometimes have failed to find a clear relationship between outcomes and length of exposure (e.g., Li et al., 2011). In the current study, very high levels of exposure that far exceeded the length of exposure prescribed by the intervention's design produced negative effects. The seemly counterintuitive result suggests that crucial information regarding the quality of implementation might have been missed in measurement and the analysis.

Investigating the quality of implementation, including the validity of self-reports on implementation, may be a key in exploring some of the emerging questions in the field. For example, why do discrepancies exist between the findings from small-scale efficacy studies and the findings from large-scale effectiveness studies (Eisner, 2009; Fraser et al., 2009)? What factors lead to discrepancies between studies led by the program developers and studies led by independent investigators (Malti, Ribeaud, & Eisner, 2011; Petrosino & Soydan, 2005)?

A variety of potential factors might explain such discrepancies. The discrepancies might stem from biases that favor the program developer, such as self-selection and expectancy effects (Malti et al., 2011). However, and perhaps more likely, the discrepancies might be due to variation in the level of administrative control over all aspects of the study. In a small-scale trail, investigators are often able to manage all aspects of the study (Petrosino & Soydan, 2005). They often have more resources to provide training for intervention agents, more frequent communication between program developers and agents, and better opportunity for negotiating resources and time for implementation. In addition, investigators in small-scale studies are more likely to have control over the selection of intervention agents (i.e., program specialists vs. classroom teachers). For example, in previous studies of the *Making Choices* program (Fraser et al., 2004; Fraser et al., 2005; Smokowski et al., 2004), the program was implemented by program specialists who were former teachers, school counselors, school psychologists, and school social workers and who were directly supervised by Fraser and his colleagues. Perhaps, a key difference between program specialists and classroom teachers is that teachers are often under additional stress caused by preparing for end-of-grade exams and mandated reforms, especially in schools that fail to meet Adequate Yearly Progress benchmarks (Fraser et al., 2009). Such pressure could substantially affect teacher engagement and implementation. This argues for detailed measurement of the quality of implementation.

Difficulties in achieving fidelity are widely reported in field settings (e.g., Gottfredson & Gottfredson, 2002). To ensure implementation fidelity, investigators have been using strategies such as training the trainers, fidelity reports, routine consultation with intervention agents, and direct observation of teacher instruction (e.g., CPPRG, 1999; Fraser et al., 2009). However, these efforts might not be able to counteract the effects of the pressure from standardized performance mandates and end-of-grade exams. Given these and other challenges, improved effects might be observed from social-emotional skills training programs under two scenarios: in one scenario the schools and teachers might held accountable for social-emotional outcomes as they are for academic performance; in the second scenario, the schools and teachers might experience less teach-to-the test pressure from the end-of-grade exams. The second scenario might be more desirable.

Although the second scenario might be more desirable, it begs the question, “Is it possible to reduce the pressure from the emphasis on academic performance?” There might be no direct and simple answer to this question. In the context of globalization that has created ever-increasing intensity of worldwide economic competition, the pressure for academic performance is ever increasing. However, it is important to understand that the enormous pressure for academic performance stems from the perception that academic performance is a principal predictor of life course success. Although there is increasing evidence indicating the importance of social-emotional skills in relation to developmental and life course outcomes (Bandura, 1999; Dodge, 1980, 2006; Huesmann, 1988; Zins, Weissberg, Wang, & Walberg, 2004; Heckman & Kautz, 2012), policymakers have yet to embrace these findings with the same degree of commitment and urgency that characterize attempts to promote academic achievement.

Limitations

To better interpret the results of the current study, several limitations must be considered. First, the loss of sample size was substantial due to missing data and the application of the common support region. The final analytic sample was about half of the original sample size. Although generalizability is not a central concern in dosage analysis, the relatively small sample resulted in a small number of observations for computing the average dosage effects at the final step of the analysis. Perhaps more important, the analysis was not able to control for bias due to differential implementation quality. Variability in implementation quality may explain, at least in part, the differences in dosage effects for groups with near-similar dosage levels.

Conclusion

Findings from this dosage analysis suggest that when programs are implemented to scale, positive effects emerge. The data here also suggest that teachers are highly variable in their implementation of SACD programs, and point to a critical area for future investigation—the quality of implementation. Implementing social-emotional skills training programs to promote social competence and prevent aggressive behavior is challenging in settings where overwhelming emphasis has been put on academic performance on standardized end-of-grade tests. The extraordinary pressure presented by end-of-grade exams and mandated reforms in schools that fail to meet the performance benchmarks of Adequate Yearly Progress is likely to have affected the acceptance and investment of teachers who delivered the program, which in turn, is likely to have affected the quality of implementation. Although implementation quality is a crucial issue in intervention research, few studies have collected data on the qualitative aspects of implementation. Investigating the quality of implementation represents an important area for future study.

Over the past 20 years, findings from rigorous studies and systematic reviews suggest that social-emotional skills are as important as academic performance in determining life course outcomes (e.g., Borghans et al., 2008; Heckman & Kautz, 2012; Heckman et al., 2006). Social-emotional skills training is increasingly recognized as an important element of the curricula of elementary schools. For example, across the 50 states, many of them have passed legislation that supports social and character development education. However, none has implemented end-of-grade testing related to this content. Although many programs can be improved and refined by incorporating new findings, seeking new strategies to ensure higher implementation quality looms large as a challenge for U.S. public schools.

REFERENCES: PAPER III

Abadie, A., & Imbens, G. (2002). *Simple and bias-corrected matching estimators for average treatment effects.* Retrieved from http://www.nber.org/papers/t0283.pdf

Arsenio, W. F., & Lemerise, E. A. (2010). Introduction: An integrative approach to emotions, aggression, and morality. In W. F. Arsenio & E. A. Lemerise (Eds.), *Emotions, aggression, and morality in children* (pp. 3-9). Washington, DC: American Psychological Association.

Baldry, A. C., & Farrington, D. P. (2005). Protective factors as moderators of risk factors in adolescence bullying. *Social Psychology of Education*, *8*, 263-284. doi:10.1007/s11218-005-5866-5

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*, 1175-1184. doi:10.1037//0003-066X.44.9.1175

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*, 193-209. doi:10.1207/s15327957pspr0303_3

Battistich, V., Schaps, E., Watson, M., & Solomon, D. (1996). Prevention effects of the Child Development Project: Early findings from an ongoing multisite demonstration trial. *Journal of Adolescent Research*, *11*, 12-35. doi:10.1177/0743554896111003

Bierman, K. L., & Wargo, J. B. (1995). Predicting the longitudinal course associated with aggressive–rejected, aggressive (non-rejected), and rejected (nonaggressive) status. *Development and Psychopathology*, *7*, 669-682. doi:10.1017/S0954579400006775

Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). *The economics and psychology of personality traits.* Retrieved from http://www.econstor.eu/bitstream/10419/34948/1/559866917.pdf

Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E.M. (1990). Preventing adolescent drug abuse through a multimodal cognitive–behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology, 58*, 437-446. doi:10.1037//0022-006X.58.4.437

Brendgen, M., Boivin, M., Vitaro, F., Bukowski, W. M., Dionne, G., Tremblay, R. E., & Perusse, D. (2008). Linkages between children's and their friends' social and physical aggression: Evidence for a gene-environment interaction? *Child Development*, 79, 13-29. doi:10.1111/j.1467-8624.2007.01108.x

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children*, *7*, 55-71. doi:10.2307/1602387

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Ferguson, L. L., & Gariepy, J.-L. (1989).Growth and aggression: 1. Childhood to early adolescence. *Developmental Psychology*, *25*, 320-330. doi:10.1037//0012-1649.25.2.320

Cairns, R. B., Leung, M.-C., Gest, S. D., & Cairns, B. D. (1995). A brief method for assessing social development: Structure, reliability, stability, and developmental validity of the Interpersonal Competence Scale. *Behavioral Research and Therapy, 33,* 725-736. doi:10.1016/0005-7967(95)00004-H

Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, *79,* 1185-1229. doi:10.1111/j.1467-8624.2008.01184.x

Coie, J. D., & Kupersmidt, J. B. (1983). A behavioral analysis of emerging social status in boys' playgroups. *Child Development*, *54,* 1400-1416. doi:10.2307/1129803

Conduct Problems Prevention Research Group (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*, 648-657. doi:10.1037/0022-006X.67.5.648

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101. doi:10.1037/0033-2909.115.1.74

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710-722. doi:10.2307/1131945

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23-45. doi:10.1016/S0272-7358(97)00043-3

Day, S. H. (2004). *Skill level activity* (Making Choices Project Tech. Rep. No. 1). Chapel Hill: University of North Carolina, School of Social Work.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51,* 162-170. doi:10.2307/1129603

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *Cognitive perspectives on children's social and behavioral development* (pp. 77-125). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dodge, K. A. (1991). Emotion and social information processing. In J. Garber & K. A. Dodge (Eds.), *The development of emotion regulation and dysregulation* (pp. 159-181). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511663963.009

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and Psychopathology ,18,* 791-814. doi:10.1017/S0954579406060391

Dodge, K. A., & Newman, J. P. (1981). Biased decision making processes in aggressive boys. *Journal of Abnormal Psychology, 90,* 375-379. doi:10.1037//0021-843X.90.4.375

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39*, 349-371. doi:10.1037//0012-1649.39.2.349

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics, 49*, 1231-1236. doi:10.2307/2532266

Eddy, J. M., Leve, L. D., & Fagot, B. I. (2001). Coercive family processes: A replication and extension of Patterson's coercion model. *Aggressive Behavior, 27*, 14-25. doi:10.1002/1098-2337(20010101/31)27:1<14::AID-AB2>3.0.CO;2-2

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology, 5*, 163-183. doi:10.1007/s11292-009-9071-y

Farrington, D. P. (1998). Evaluating "communities that care": Realistic scientific considerations. *Evaluation, 4*, 204-210. doi:10.1177/13563899822208536

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology, 36*, 127-151. doi:10.1375/acri.36.2.127

Feldman, E., & Dodge, K. A. (1987). Social information processing and sociometric status: Sex, age, and situational effects. *Journal of Abnormal Child Psychology, 15,* 211-227. doi:10.1007/BF00916350

Flannery, D. J., Vazsonyi, A. T., Liau, A. K., Guo, S., Powell, K. E., Atha, H., Vesterdal, W., & Embry, D. (2003). Initial behavior outcomes for the PeaceBuilders universal school-based violence prevention program. *Developmental Psychology, 39*, 292-308. doi:10.1037/0012-1649.39.2.292

Flavell, J. H., Botvin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children*. New York, NY: Wiley.

Flores, C. A., Flores-Lagunes, A., Gonzalez, A., & Neumann, T. C. (2010). *Estimating the effects of length of exposure to instruction in a training program: The case of Job Corps.* Retrieved from http://moya.bus.miami.edu/~cflores/research_papers/FNGF-DRF-031610.pdf

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

Fraser, M. W. (2004). Intervention research in social work: Recent advances and continuing challenges. *Research on Social Work Practice, 14,* 210-222. doi:10.1177/1049731503262150

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. doi:10.1177/1049731503257884

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. Y. (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,* 1045-1055. doi:10.1037/0022-006X.73.6.1045

Fraser, M. W., Kirby, L. D., & Smokowski, P. R. (2004). Risk and resilience in childhood. In M. W. Fraser (Ed.), *Risk and resilience in childhood: An ecological perspective* (2nd ed., pp. 13-66). Washington, DC: NASW Press.

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

Fraser, M. W., Guo, S. Y., Ellis, A. R., Day, S. H., Li, J. L., Wike, T. L., & Farmer, T. W. (2009). *Social and character development in elementary school: Effects from a controlled trial*. (Report to Institute of Education Sciences and the Centers for Disease Control and Prevention). Retrieved from http://sowkweb.usc.edu/research/events/papers/Social-and-Character-Development-in-Elementary-School.pdf

Fraser, M. W., Guo, S. Y., Ellis, A., R., Thompson, A. M., Wike, T. L., & Li., J. L. (2011). Outcome studies of social, behavioral, and educational interventions: Emerging issues and challenges. *Research on Social Work Practice, 21*, 619-635. doi:10.1177/1049731511406136

Gouze, K. R. (1987). Attention and social problem solving as correlates of aggression in preschool males. *Journal of Abnormal Child Psychology, 15*, 181-197. doi:10.1007/BF00916348

Gresham,F. M., Gansle,K. A., Noell, G. H., Cohen, S., & Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review*, *22*, 254-272.

Grossman, D., Neckerman, H. J., Koepsell, T. D., Liu, P., Asher, K., Beland, K., Frey, K., & Rivara, F. P. (1997). Effectiveness of a violence prevention curriculum among children in elementary school: A randomized controlled trial. *Journal of the American Medical Association, 277*, 1605-1611. doi:10.1001/jama.277.20.1605

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency, 39*, 3-35. doi:10.1177/002242780203900101

Grotpeter, J. K., & Crick, N. R. (1996). Relational aggression, overt aggression, and friendship. *Child Development*, *67*, 2328-2338. doi:10.2307/1131626

Guerra, V. S., Asher, S. R., DeRosier, M. E. (2004). Effect of children's perceived rejection on physical aggression. *Journal of Abnormal Child Psychology, 32*, 551-563. doi:10.1023/B:JACP.0000037783.88097.69

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage.

Guo, S., & Hussey, D. (1999). Analyzing longitudinal rating data: A three-level hierarchical linear model. *Social Work Research*, *23*(4), 258-269. Retrieved from http://swr.oxfordjournals.org/content/23/4/258.full.pdf

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., …Dahlberg, L. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine*, *33*, S114-S129. doi:10.1016/j.amepre.2007.04.012

Hanish, L. D., & Guerra, N. G. (2002). A longitudinal analysis of patterns of adjustment following peer victimization. *Development and Psychopathology, 14*, 69-89. doi:10.1017/S0954579402001049

Heckman, J. J., & Kautz, T. D. (2012). *Hard evidence on soft skills.* Retrieved from http://www.nber.org/papers/w18121.pdf?new_window=1

Henry, D.B., Tolan, P.H., Gorman-Smith, D. (2001). Longitudinal family and peer groupeffects on violence and nonviolent delinquency. *Journal of Clinical Child Psychology*, *30*, 72-186. doi:10.1207/S15374424JCCP3002_5

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). *The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.* Retrieved from http://www.nber.org/papers/w12006.pdf?new_window=1

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete-data perspectives* (pp.73-84). New York, NY*:* John Wiley. doi:10.1002/0470090456.ch7

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

Imbens, G. W. (2000). The role of propensity score in estimating dose-response functions. *Biometrika, 87*, 706-710. doi:10.1093/biomet/87.3.706

Jenson, J. M., & Howard, M. O. (1999). *Youth violence: Current research and recent practice innovations.* Washington, DC: NASW Press.

Jenson, J. M., & Howard, M. O. (2001). Causes and prevention of youth violence. *Denver University Law Review, 77*, 629-660.

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of Epidemiology, 150*, 327-333. doi:10.1093/oxfordjournals.aje.a010011

Lechner, M., & Melly, B. (2010). partial identification of wage effects of training programs. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1596715

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. doi:10.1016/j.appdev.2003.08.002

Li, K. K., Washburn, I., DuBois, D. L., Vuchinich, S., Ji, P., Brechling, V., … Flay, B. R. (2011). Effects of the Positive Action programme on problem behaviours in elementary school students: A matched-pair randomised control trial in Chicago. *Psychology & Health*, *26*, 187-204. doi:10.1080/08870446.2011.531574

Li, J. L. (2012). *Estimating causal effects with propensity score: The essential ideas and general principles*. Unpublished manuscript.

Li, J. L., Fraser, M . W., & Wike, T., L. (in press). Promoting social competence and preventing childhood aggression: A framework for applying social information processing theory in intervention research. *Aggression and Violent Behavior*.

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. doi:10.1177/104973150201200204

Malti, T., Ribeaud, D., & Eisner, M. (2011). The effects of two universal preventive interventions to reduce children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40,* 677-692. doi:10.1080/15374416.2011.597084

McIntosh, M. W., & Rubin, D. (1999). On estimating the causal effects of DNR orders. *Medical Care, 37,* 722-726. doi:10.1097/00005650-199908000-00002

Merrell, K., Gueldner, B.A., Ross, S.W., & Isava, D.M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23*, 26-42. doi:10.1037/1045-3830.23.1.26

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Developmental Psychopathology, 13,* 355-375. doi:10.1017/S0954579401002097

Moncher, F. J., & Prinz, R. J. (1991).Treatment fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266. doi:10.1016/0272-7358(91)90103-2

Multisite Violence Prevention Project. (2009). The ecological effects of universal and selection violence prevention programs for middle school students: A randomized trial. *Journal of Consulting and Clinical Psychology, 77*, 526-542. doi:10.1037/a0014395

Nash, J. K., Fraser, M. W., Galinsky, M. J., & Kupper, L. L. (2003). Early development and pilot testing of a problem-solving skills-training program for children. *Research on Social Work Practice, 13*, 432-450. doi:10.1177/1049731503013004002

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N. P., Hancox, R., Harrington, H., ... Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology, 20*, 673-716. doi:10.1017/S0954579408000333

Park-Higgerson, H. K., Perumean-Chaney, S. E., Bartolucci, A. A., Grimley, D. M., & Singh, K. P. (2008). The evaluation of school-based violence prevention programs: A meta-analysis. *Journal of School Health, 78*,465-479. doi:10.1111/j.1746-1561.2008.00332.x

Patterson, J. M. (2002). Understanding family resilience. *Journal of Clinical Psychology, 58,* 233-246. doi:10.1002/jclp.10019

Payton, J. W., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M (2008). *Positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews.* Retrieved from the Collaborative for Academic, Social, and Emotional Learning website. Available at: http://casel.org/wp-content/uploads/PackardTR.pdf

Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Anderson Johnson, C., …Cormack, C. (1990). Effects of program implementation on adolescent drug use behavior: The Midwestern Prevention Project. *Evaluation Review, 14*, 264-289. doi:10.1177/0193841X9001400303

Perkins, S. M., Tu, W., Underhill, M. G., Zhou, X.-H., & Murray, M. D. (2000). The use of propensity scores in pharmacoepidemiologic research. *Pharmacoepidemiology and Drug Safety, 9*, 93-101. doi:10.1002/(SICI)1099-1557(200003/04)9:2<93::AID-PDS474>3.0.CO;2-I

Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology, 1*, 435-450. doi:10.1007/s11292-005-3540-8

Pettit, G. S., Dodge, K. A., & Brown, M. M. (1988). Early family experience, social problem solving patterns, and children's social competence. *Child Development, 59,* 107-120. doi:10.1111/j.1467-8624.1988.tb03199.x

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. doi:10.1037/0022-006X.72.1.103

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55. doi:10.1093/biomet/70.1.41

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, 757-763.

Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health*, *38,* 32-43. doi:10.1007/s10488-010-0321-0

Selman, R. L. (1971). Taking another's perspective: Role-taking development in early childhood. *Child Development, 42,* 1721-1734. doi:10.2307/1127580

Shantz, C. U. (1975). The development of social cognition. In E. M. Hetherington (Ed.), *Review of child development research* (Vol. 5, pp. 257-323). Chicago, IL: University of Chicago Press.

Shantz, C. U. (1983). Social cognition. In J. H. Flavell & E. M. Markman (Eds.),

Handbook of child psychology: Vol. 3. Cognitive development (4th ed., p. 495-555). New York, NY: John Wiley & Sons.

Smokowski, P. R., Fraser, M. W., Day, S. H., Galinsky, M. J., & Bacallao, M. L. (2004). School-based skills training to prevent childhood aggression: Using the Making Choices program as a universal prevention initiative. *Journal of Primary Prevention, 25,* 233-251. doi:10.1023/B:JOPP.0000042392.57611.05

Thornberry, T. P., Huizinga, D., & Loeber, R. (2004). The causes and correlates studies: Findings and policy implications. *Juvenile Justice, 9,* 3-19.

Underwood, M. K. (2003). *Social aggression among girls*. New York, NY: Guilford Press.

Wasserman, G. A., & Seracini, A. G. (2001). Family risk factors and interventions. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention, and service needs* (pp. 165-189). Thousand Oaks, CA: Sage.

Waters, E., & Sroufe, L. A. (1983). Social competence as a developmental construct. *Developmental Review, 3*, 79-97. doi:10.1016/0273-2297(83)90010-2

Weissberg, R. P., & Elias, M. J. (1993). Enhancing young people's social competence and health behavior: An important challenge for educators, scientists, policy makers, and funders. *Applied & Preventive Psychology: Current Scientific Perspectives*, *3*, 179-190. doi:10.1016/S0962-1849(05)80088-5

Weissberg, R. P., & Greenberg, M. T. (1998). School and community competence-enhancement and prevention programs. In I. E. Siegel & K. A. Renninger (Vol. Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 877-954). New York, NY: Wiley.

Williams, J. H., Ayers, C. D., Van Dorn, R. A., & Arthur, M. W. (2004). Risk and protective factors in the development of delinquency and conduct disorder. In M. W. Fraser (Ed.), *Risk and resiliency in childhood: An ecological perspective* (2nd ed., pp. 209-249). Washington, DC: NASW Press.

Williamson, E., Morley, R., Lucas, A., & Carpenter, J. (2012). Propensity scores: from naive enthusiasm to intuitive understanding. *Statistical Methods in Medical Research, 21*, 273-293. doi:10.1177/0962280210394483

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. doi:10.1037/0022-006X.71.1.136

Wilson, S. J., & Lipsey, M. W. (2006). *The effects of school-based social information processing interventions on aggressive behavior. Part I: Universal programs.* (Campbell Systematic Review 2006:5). doi: 10.4073/csr.2006.5 Retrieved from http://campbellcollaboration.org/lib/download/68/

Wilson, S. J. & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine, 33*, S130-S143. doi:10.1016/j.amepre.2007.04.011

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics, 20*(1), 59-73. doi:10.3102/10769986030001059

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building academic success on social and emotional learning: What does the research say?* New York, NY: Teachers College Press.

SUMMARY

Social-emotional skills developed in childhood are associated with negative developmental outcomes such as peer rejection and aggressive behavior (e.g., Smith, 2001; Trentacosta & Izard, 2007). In addition, they are related to long-term academic, mental health, and socioeconomic outcomes (e.g., Heckman & Kautz, 2012). Therefore, promoting social-emotional skills and preventing aggressive behavior in childhood are crucial areas for interventions.

Social-emotional skills training is increasingly recognized as an important element of the curricula of elementary schools. However, evidence supporting the effectiveness of social-emotional skills training programs has not been well-established. Although, systematic reviews of evaluation studies of universal school-based social-emotional skills training programs suggest that the majority of these programs was effective (e.g., Farrington & Welsh, 2003; Payton et al., 2008; Wilson & Lipsey, 2006), the effect sizes are small in general. Moreover, a number of studies have reported contradictory program findings (e.g., CPPRG, 1999; Eisner, 2009; Flannery et al., 2003; Institute of Education Sciences, 2011; Malti, Ribeaud, & Eisner, 2011; Merrell, Gueldner, Ross, & Isava, 2008).

The lack of strong evidence supporting the effectiveness of social-emotional skills training warrants continuing efforts to refine existing programs or design new programs by incorporating new knowledge from basic social behavior sciences. The development of the SIP model represents a substantial advance in understanding how social cognition affects the behavioral responses of children in social interactions. By specifying cognitive operations

underlying the behaviors of children, SIP theory has important applications for designing interventions to improve social- emotional skills and prevent aggressive behavior in childhood.

However, the translation of the SIP model to social-emotional skills training is still in a formative stage; only a few programs have explicitly applied SIP theory. Moreover, applications of the SIP model vary greatly across programs (e.g., Fraser et al., 2005; Frey et al., 2000; Meyer & Farrell, 1998). Most applications are characterized by focusing on selected SIP steps rather than all of the sequential steps (e.g., Frey et al., 2000; Meyer & Farrell, 1998; Sawyer et al., 1997; Work & Olsen, 1990). Essentially, the SIP model has been used to modify training strategies within a traditional SPS framework.

The existing social-emotional training programs vary widely in their theoretical foundations and the activities that carry out theory. Two decades ago, Ladd and Mize (1982) called for a precise and unified model of social-skills training. Developing a precise and unified model for social-emotional skills training programs remains a challenge to intervention researchers. Promoting communication and collaboration between intervention researchers in multiple disciplines is critical in developing more effective skills-training programs.

In promoting social-emotional skills training, researchers also need to address issues presented by contradictory program findings. Understanding factors that contribute to the mixed findings is crucial. Dosage analysis has the potential to untangle program effects from effects due to variation in implementation. Findings from dosage analyses also provide crucial information regarding optimal exposure to (or dose of) an intervention.

Despite the utility of dosage analyses, such analyses remain an understudied area due perhaps primarily to the emphasis on intent-to-treat analysis. Recently, researchers have called for attention to the consequence of conflating program effects with implementation effects and have reemphasized the importance of dosage analysis (Fraser et al., 2011). To be sure, conducting a dosage analysis is challenging because it often requires balancing multiple groups simultaneously—a task that is typically beyond the capacity of conventional regression methods. GPS methods provide a viable means for balancing groups defined by different dosages. GPS methods are a recent development in the family of propensity score methods. The introduction of these methods to intervention researchers is expected to facilitate efforts in untangling program effects from effects of varying implementation.

Undoubtedly, the successful application of GPS methods is contingent on the plausibility of statistical assumptions. GPS methods require a weak version of the unconfoundedness assumption. The key implication of this unconfoundedness assumption is the absence of unmeasured confounders. Thus, a successful use of GPS methods requires prudence in identifying and measuring confounders before embarking on the analysis. Any important confounders left unmeasured would have the potential to result in a biased estimation of program effects.

Applying the GPS method with continuous treatment, this dissertation study investigated dosage effects in a SIP-based program, *Making Choices*. Findings from the dosage analysis suggest that intervention effects vary by treatment exposure. Moreover, variation of dosage effects for participants with similar exposure points to a critical area for future investigation—that is, the quality of implementation.

Assuring implementation quality is challenging in settings where overwhelming emphasis has been put on academic performance on standardized end-of-grade tests. The extraordinary pressure presented by end-of-grade exams and mandated reforms in schools that fail to meet the performance benchmarks of adequate yearly progress is likely to have affected the acceptance and investment of teachers who delivered the program, which in turn, is likely to have affected the quality of implementation.

Over the past 20 years, findings from rigorous studies and systematic reviews suggest that social-emotional skills are as important as academic performance in determining life-course outcomes (e.g., Borghans et al., 2008; Heckman & Kautz, 2012). However, these findings have not been embraced by policymakers and educators with the same degree of commitment and urgency that characterize attempts to promote academic achievement. Although social emotional-skills training is increasingly recognized as an important element of the curricula of elementary schools, no school has implemented end-of-grade testing related to this content.

In summary, the critical role of social-emotional skills in predicting life-course outcomes warrants continuing efforts to refine and develop new intervention programs by incorporating new findings from social-behavioral research. In this effort, the communication and collaboration of researchers in multiple disciplines are critical for the field to develop a precise and unified program of social-emotional skills training. Meanwhile, seeking strategies to assure higher implementation quality looms large as a challenge for U.S. public schools.

REFERENCES: SUMMARY

Borghans, L., Duckworth, A. L., Heckman, J. J., & Weel, B. T. (2008). The economics and psychology of personality traits. Retrieved from http://www.econstor.eu/bitstream/10419/34948/1/559866917.pdf

Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67,* 648-657. doi:10.1037/0022-006X.67.5.648

Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, *5*, 163-183. doi:10.1007/s11292-009-9071-y

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology, 36,* 127-151. doi:10.1375/acri.36.2.127

Flannery, D. J., Vazsonyi, A. T., Liau, A. K., Guo, S., Powell, K. E., Atha, H., …Embry, D. (2003). Initial behavior outcomes for the PeaceBuilders universal school-based violence prevention program. *Developmental Psychology, 39,* 292-308. doi:10.1037/0012-1649.39.2.292

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. Y. (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,* 1045-1055. doi:10.1037/0022-006X.73.6.1045

Fraser, M. W., Guo, S. Y., Ellis, A., R., Thompson, A. M., Wike, T. L., & Li., J. L. (2011). Outcome studies of social, behavioral, and educational interventions: Emerging issues and challenges. *Research on Social Work Practice, 21*, 619-635. doi:10.1177/1049731511406136

Frey, K. S., Hirschstein, M. K., & Guzzo, B. A. (2000). Preventing aggression by promoting social competence. *Journal of Emotional and Behavioral Disorders, 8,* 102-112. doi:10.1177/106342660000800206

Heckman, J. J., & Kautz, T. D. (2012). *Hard evidence on soft skills*. Retrieved from http://www.nber.org/papers/w18121.pdf?new_window=1

Institute of Education Sciences. (2011). *Efficacy of schoolwide programs to promote social and character development and reduce problem behavior in elementary school children.* Retrieved from http://ies.ed.gov/ncer/pubs/20112001/index.asp

Ladd, G. W., & Mize, J. (1983). A cognitive-social learning model of social skill training. *Psychology Review, 90,* 127-157. doi:10.1037/0033-295X.90.2.127

Malti, T., Ribeaud, D., & Eisner, M. (2011). The effects of two universal preventive interventions to reduce children's externalizing behavior: A cluster randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology, 40*, 677-692. doi:10.1080/15374416.2011.597084

Merrell, K., Gueldner, B.A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly*, *23*, 26-42. doi:10.1037/1045-3830.23.1.26

Meyer, A. L., & Farrell, A. D. (1998). Social skills training to promote resilience in urban sixth-grade students: One product of an action research strategy to prevent youth violence in high-risk environments. *Education and Treatment of Children, 21*, 461-488.

Payton, J. W., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M (2008). *Positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews.* Retrieved from http://casel.org/wp-content/uploads/PackardTR.pdf

Smith, M. (2001). Social and emotional competencies: Contributions to young African-American children's peer acceptance. *Early Education and Development, 12*, 49-72. doi:10.1207/s15566935eed1201_4

Sawyer, M. G., Macmullin, C., Graetz, B., Said, J. A., Clark, J. J., & Baghurst, P. (1997). Social skills training for primary school children: A 1-year follow-up study. *Journal of Paediatrics and Child Health, 33,* 378-383. doi:10.1111/j.1440-1754.1997.tb01624.x

Trentacosta, C. J., & Izard, C. E. (2007). Kindergarten children's emotion competence as a predictor of their academic competence in first grade. *Emotion, 7*, 77-88. doi:10.1037/1528-3542.7.1.77

Wilson, S. J., & Lipsey, M. W. (2006). *The effects of school-based social information processing interventions on aggressive behavior. Part I: Universal programs*. (Campbell Systematic Review 2006:5). doi:10.4073/csr.2006.5 Retrieved from http://campbellcollaboration.org/lib/download/68/

Work, W. C., & Olsen, K. H. (1990). Evaluation of a revised fourth grade social problem-solving curriculum: Empathy as a moderator of adjustive gain. *Journal of Primary Prevention, 11*, 143-157.