5-13-2011

A Multi-Level Investigation of Teacher Instructional Practices and the Use of Responsive Classroom

Benjamin George Solomon
University of Massachusetts Amherst, bgsolomo@educ.umass.edu

A MULTI-LEVEL INVESTIGATION OF TEACHER INSTRUCTIONAL PRACTICES AND THE USE OF RESPONSIVE CLASSROOM

A Dissertation Presented

by

BENJAMIN G. SOLOMON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 2011

School Psychology

A MULTI-LEVEL INVESTIGATION OF TEACHER INSTRUCTIONAL PRACTICES AND THE USE OF RESPONSIVE CLASSROOM

A Dissertation Presented

by

BENJAMIN G. SOLOMON

Approved as to style and content by:

_____
John M. Hintze, Chairperson

_____
Amanda M. Marcotte, Member

_____
David G. Scherer, Member

_____
Christine B. McCormick, Dean
School of Education

DEDICATION

To my wife, Emily, who provided endless support and whose love of teaching inspired this project.

ACKNOWLEDGEMENTS

I am extremely grateful to those who invested in my training and helped shape the current study. Thanks to John Hintze, my advisor and committee chair. Your mentorship and dedication to empirical inquiry inspired in me a lifelong commitment to promote evidence-based practice. Thank you to committee member Amanda Marcotte, whose practical understanding of classroom ecology was essential. Thanks to committee member David Scherer, whose objective yet supportive feedback broadened my understanding of systems level research. I would like to recognize the assistance of Craig Wells, whose support was indispensable in calculating results. I also would like to express my utmost appreciation to the Wing Institute of Oakland, CA, who funded this research.

This project was not possible without the expertise, friendship and enthusiasm of my research partner, Suzanne Klein. It is remarkable to reflect on how, from a conversation, emerged such a project. I wish to acknowledge the contributions of my peers James Cressey and Sarah Peller. Their understanding of prevention and research methods helped guide the early development of this project. Thank you to the research observers: Katie Meyer, Abi Leibovitch, Rebecca Allen, Kara White, Michelle Heller, Danille Lucey, and Cheyne Levessuer.

Finally, thank you to all the teachers and administrators that generously gave their time and feedback to allow this study to happen. In particular, I would like to recognize Melinda Foley-Marsello of the Friends Academy, Rose Mary Grant of the CVS Highlander School, Barbara Malkas and Kevin Tobin of the Pittsfield Public School District, and Philip DiPietro of the Hadley Elementary School.

ABSTRACT

A MULTI-LEVEL INVESTIGATION OF TEACHER INSTRUCTIONAL PRACTICES AND THE USE OF RESPONSIVE CLASSROOM

MAY 2011

BENJAMIN G. SOLOMON, B.A., MUHLENBERG COLLEGE

M.Ed, UNIVERSITY OF MASSACHUSSETTS AMHERST

Ph.D., UNIVERSTY OF MASSACHUSETTS AMHERST

Directed by: Professor John M. Hintze

A year-long longitudinal study was conducted to quantify different types of teaching in the beginning of the year, and the effect of those choices on end of year instructional practices and student outcomes. Teacher practices were organized around the fidelity of implementation to the *Responsive Classroom* (*RC*) program (Northeast Foundation for Children, 2009). Most notably, a central *RC* tenant entitled "the first six weeks" was examined. *RC* is a universal prevention program that previously has been categorized as a Tier I social-behavioral program for students when considered within an RTI model (Elliott, 1999).

Twenty-seven teachers from the New England region and 179 students participated. The Academic Competence Evaluation Scales (ACES), teacher-form (DiPerna & Elliott, 2000) was used to measure student outcomes. The Classroom Practice Measure (CPM; Rimm-Kaufman et al., 2007) was used to measure level of *RC* implementation. Finally, to quantify teaching behavior, a momentary time-sampling observation, called the Teaching Observation Tool (TOT; Marcotte, Klein, & Solomon, 2010), was implemented.

Results from a series of multilevel models utilizing students nested within teachers indicated that both a constant, high level of instructional time and investment in environmental management time in the fall results in higher levels of student reading (significant) and math achievement (non-significant) in the spring, and lower levels of time spent correcting behavior. Teachers with large discrepancies in instructional time from fall to spring and teachers who failed to release environmental control to students over time had students with lower levels of reading and math growth.

Relationships between the CPM, ACES, and the TOT indicate that RC is significantly correlated with increases in student reading achievement and motivation beyond what would be expected of a teacher that does not implement RC. However, in contrast to past research, *RC* in this study was not correlated with teacher reported improvements in social skills. Implications for practice and directions for future research are discussed.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

CHAPTER I

PROBLEM STATEMENT AND LITERATURE REVIEW

Statement of the Problem

The need for school-wide social-behavioral instruction has drawn increased attention in the past decade from popular media, federally funded empirical studies (cf. Gottfredson, Gottfredson, Czeh, Cantor, Crosse, & Westat, 2000; Nansel, Overpeck, Pila, Ruan, Simons-Morton, & Schedit 2001; Satcher, 2001) and contemporary research in the field. Facilitating positive social behavior and preventing negative, antisocial behavior can serve as an academic enabler by allowing students more instructional time and enriching interactions between students (Elliott, DiPerna, Mroch, & Lang, 2004; Vygotsky, 1978). A comprehensive school-wide social-behavioral prevention system, implemented immediately upon school matriculation, can inoculate students against mental health problems that otherwise would increase in severity over time (Ford & Lerner, 1992).

Social skills also represent a set of discrete behaviors that are critical to a student's overall adaptive functioning and readiness for post-secondary education and employment. The necessity for instruction in social-behavior skills only exacerbates the need for evidence-based prevention in the field. Research needs to be directed at demonstrating both the efficacy and effectiveness of prevention and intervention programs across a variety of contexts to adequately address this need. Although theoretically prevention services could be delivered outside schools - schools allow a finer control over the environment, allow rapid generalization of skills, and permit

frequent opportunities to practice skills with other students (Payne, Gottfredson, & Gottfredson, 2006).

Satcher (2001) reported that cases of documented aggravated assault and robbery involving youths has escalated at a staggering rate; nearly 70% from 1980 to 1999, despite other forms of criminal activity decreasing in past years. Hoagwood and Erwin (1997) reported that up to 22% of students enrolled in public education have behavioral problems severe enough to warrant mental health services. Many of these students go undetected, leading to multiple problems that increase in cost and severity and decrease in chance of remediation over time (Sprague & Walker, 2005). The expression of these problems in the classroom takes away from instructional time, even at mild levels of severity, and can give rise to a culture of antisocial behavior within schools that inhibits both academic and social-emotional growth.

Gottfredson et al. (2000) sampled 6,451 schools by surveying teachers, principals, and students in regards to problems with delinquent behavior and prevention efforts. 6.7% of school principals reported more than one incident of severe physical aggression in their school annually. Gottfredson et al. (2000) noted that principal report was only modestly correlated to other prevalence statistics. When students were asked directly, rates of student involvement either as the victim or perpetrator of aggression was as high as 41%, demonstrating that most school-based antisocial behavior is covert and goes undetected. Furthermore, up to 27% of teachers surveyed believed externalizing student behavior present in their classrooms prevented them from delivering effective instruction. In comparison Nansel et al. (2001), in a national sample of 15,686 students in middle and high school, found that 30% of students surveyed were involved in bullying, either as

victims, bullies, or both. Prevalence rates were consistent across demographic region (e.g., urban, surburban, and rural). Given that schools are vulnerable to becoming grounds for student victimization, they become prime locations to deliver intervention.

Using data from the 1998 National Survey of Drug Use and Health, Miller (2004) performed a cost-effectiveness analysis on the outcomes of adolescents who engage in violence, binge drinking, drug use, dangerous sexual behavior, suicide, and school dropout. Miller (2004) summed combined medical costs, resource costs, work costs, and quality of life costs (a monetary measure of suffering and degeneration of life quality incurred due to dangerous behavior) using this data set. He estimated that up $334 billion were spent or lost across all categories on multi-problem adolescents. A separate estimate showed that violent behavior accounted for $165 billion of lost income and incurred expenses and was the most expensive adolescent problem. High school dropout was second and accounted for $141 billion. While Miller's (2004) estimates were based on national averages, such problems likely draw heavily on school resources and teacher's classroom management resources as well. This is particularly salient considering many schools are currently under significant fiscal strain.

Evidence-based prevention efforts, when implemented early in a student's development, improves the outcome of the student both behaviorally and academically, and in doing so improves the overall health and functioning of the school. As Elliott, Hamburg and Williams (1998) summarize, there is a longstanding counterintuitive notion held that reactive intervention is more cost-effective than frontloading prevention efforts. The distribution of funds to reactive services such as state correctional facilities or residential services for students far outweighs the funding given to schools for research,

implementation and maintenance to prevent student problems when the severity of symptoms are low and are more easily remediated. The potential benefits of such prevention programming are far from trivial. Dodge and Sherrill (2007) summarized 30 studies on the transactional effect of environment and person-based variables that leads to childhood developmental violence. These attributes ranged from levels of the MAO-A gene, to birth complications, to personality risk factors. They found across studies that genetic expression was modulated by environmental attributes. Children who were most at-risk for a violent predisposition would only express such violent tendencies if the environment also was conducive to violent behavior (e.g., childhood trauma, poverty, etc.). Children at-risk for violent behavior were indistinguishable from low-risk students when the environment favored a positive, non-violent upbringing.

From this research, one could hypothesize that school-based social-behavioral prevention would ameliorate or build resistance to environmental risk-factors, negating person-based risk factors. This hypothesis has been confirmed by contemporary research. van Lier, Muthen, van der Sar, and Criine (2004) implemented the *Good Behavior Game* in 31 classrooms in the Netherlands as a form of universal prevention. The severity of conduct problems at baseline served as the measure of risk status and the primary dependent variable. After two years of intervention, conduct problems were reduced in proportion to the baseline rate of behavior: $d = .55$ for the highest at-risk, $d = .42$ for those moderately at-risk, and no significant effect for those with no reported risk factors. However, a critical factor in this remedial process is the assumption that any given program positively affects the school-based environment in which the student resides; that the program is evidence-based for the target population.

Expert rating systems have emerged recently, such as the *Collaborative for Social and Emotional Learning* (CASEL; 2009) database, or the *What Works Clearinghouse* (Institute of Educational Science, 2009), which have provided invaluable insight for consumers into the quality of various tier I programs for social skills and behavioral prevention. Unfortunately, these resources are reliant on controlled, experimental data, which are sorely lacking for many social-behavioral prevention programs. Education consumers need to have access to contemporary research that can lead to the selection of a well validated program that matches the specific needs of the local educational agency.

An Example of a Social-Behavior Prevention Program: The Responsive Classroom Approach

In the past twenty years, *Responsive Classroom* (*RC*) has proliferated as a universal social-behavioral prevention program for elementary school students. The program utilizes a set of teaching strategies and philosophies to promote positive behavior using a community-based approach. *RC* is sponsored by the Northeast Foundation for Children (NEFC), which has developed into a substantial non-profit organization, drawing revenue through in-school consultation, professional development and a wide variety of purchasable program related supplies and texts related to *RC* (NEFC, 2009). Research on school-based social-behavioral prevention has demonstrated that universal level investment can lead to significant positive outcomes for students and is more cost-effective than a reactive intervention approach alone (Biglan, Brennan, Foster, & Holder, 2004; Gresham, Sugai,& Horner, 2001; Ross, Powell, & Elias, 2002). However for such a benefit to occur, a restrictive set of assumptions based on the criteria

for evidence-based prevention and Response to Intervention (RTI) must be met (Flay et al., 2005; Fuchs & Fuchs, 2006; Gresham et al., 2001; Nation et al., 2003; Sheridan, Hungelmann, & Maughan, 1999).

Summary of the Responsive Classroom Curriculum

RC, like other comprehensive universal prevention programs, has been developed as a constellation of educational philosophies that have emerged into explicit practices over time. Rimm-Kaufman and Chiu (2007) describe *RC* as rooted in two primary theories: the Bioecological Model (Bronfenbrenner & Morris, 2006) and General Systems Theory (GST; Bertalanffy, 1968; Pianta, 1999; Sameroff, Emde, & Anders, 1989). The Bioecological Model is composed of four core variables that influence human development: the person, proximal processes, the context, and time. Rimm-Kaufman and Chiu (2007) highlight the role of proximal processes; the bidirectional influence between the unique elements of the person including biological and psychological variables and the context(s) the person is embedded in. The context can be seen as a system of overlapping uniform environmental clusters, ranging from macro-level cultural beliefs to the influences of various microsystems such as the school and home environment (Bronfenbrenner & Morris, 2006). Bronfenbrenner and Morris (2006) describe how proximal processes accrue over time to influence behavior in an adaptive or maladaptive route. In *RC*, it is believed by shifting proximal processes to positive interactions that are rich in social exchanges that child development can be nurtured. A sensitivity to the student's developmental level, unique to each student regardless of chronological age, is a hallmark of the *RC* approach (NEFC, 2003).

GST is summarized by Bertalanffy (1968), who demonstrated the universality of systems theory as an epistemology by applying it to a broad range of disciplines: from thermodynamics to psychiatry. GST is rooted in early theories of causal mechanisms in cellular biology. Bertalanffy (1968) stated that the significance of any individual component in a system, such as an atom or cog, is severely limited without an understanding of its relation to proximal units that form an overarching purpose. In this review, Bertanffy stated that GST is a framework that explains how certain conditions, such as psychopathology, are modulated by both a complex internal system and a larger social system.

GST is explained in a school-based child development context by Pianta (1999). GST bears similar resemblance to the Bioecological Model, although focuses more exclusively on the role of context. While the Bioecological Model explains context as a variable that interacts with person-based variables, GST focuses on context as having a regulatory function in development, with specific expectations embedded within each environmental system from distal to proximal (Sameroff et al., 1989). The teacher attempts to regulate behavioral and academic growth; the principal attempts to regulate school policy that affects teacher behavior; the community attempts to regulate principal decisions through a school committee and so forth. These layered regulatory bodies create a set of sociocultural expectations for students, which they may or may not be able to meet based on biological potential, past experiences, and the interaction of environments the student is nested in.

Pianta's (1999) unique contribution to the model is seen in his application of GST to general education instruction for high-risk students. Pianta stated that to keep specific

contexts stable, schools should cease progressively specializing and fragmenting services (i.e., pull-out services) as much as possible (1999). Additionally, students who are labeled at-risk should be able to build a deep bond with few individuals, suggesting teachers follow students through grade levels so teachers can regulate student development and serve several roles. *RC*, as a universal level program which aims to facilitate teacher-child relationships and prevent problems that may damper this relationship or lead to pull-out services, is in line with this theory.

Recent texts have described *RC* as a constellation of seven guiding principles based on the above mentioned theories and the individual experiences of the *RC* developers: (1) a focus on both social and academic learning; (2) a focus on the process of learning as well as the net gain; (3) a recognition of the value of social interaction in classrooms; (4) a focus on behaviors that facilitate community involvement such as cooperation and empathy; (5) a focus on understanding all the contexts a child is embedded in and recognizing that each child has unique influences; (6) involving the family and the community in the learning process; (7) and creating a professional, collaborative environment amongst teachers (Rimm-Kaufman & Chiu, 2007; Rimm-Kaufman, Fan, Chiu, & You, 2007).

Over time, these philosophies have developed into various teacher practices. While the program is designed to be comprehensive, recent evaluations have shown many teachers select specific practices to adopt out of the general program (Elliott, 1999). The most commonly used practices can be summarized as follows, based on Elliott (1999), NEFC (2003) and Rimm-Kaufman and Chiu (2007):

Use of a structured morning meeting. The teacher is expected to devote time daily to a structured morning meeting to build a classroom community, practice social skills and go over daily routines. The meeting is typically divided into four components: A greeting, sharing of personal news, a group activity, and news and announcements.

Classroom organization. Classroom arrangement of materials and furniture is setup in a thoughtful manner that allows for open group meetings, a variety of places to work, and clear labels for materials. This practice also includes prominent displays of class work and art, and avoidance of clutter.

Rules and logical consequences. After routines are established by the teacher, rules are jointly generated by both the teacher and the students, with the purpose of achieving explicit community-based goals. Rules are typically framed in the positive and reflect general values. Practice, modeling, and reflection help facilitate positive student behavior. The teacher uses specific behavioral language, framed in training as the "three R's": reinforce, remind, redirect. Violation of classroom rules leads to "logical consequences"; sanctions that are "respectful", "relevant", and "realistic". Consequences are intended to fit the situation and result in "fixing" the damage done to the social relationship between the involved community members. An example is an "apology of action", in which a student will do a specific positive action to make up for aggressive or rude behavior to another student. Time-out is endorsed when used wisely and to prevent more severe negative behavior.

Guided discovery. The teacher states lesson objectives clearly and prompts understanding of directions and proper use of materials through open-ended questions.

Students are given flexibility to creatively try out ideas within lessons and generate their own hypothesis.

Academic choice time. Center activities and small group instruction allow creative use of materials and objectives to foster student curiosity. Students are encouraged to reflect on their use of materials and their perception of tasks. Work is leveled appropriately so all students are given some locus of control over the centers and use of time.

While each of these instructional techniques target a different set of behavioral and learning objectives, in general their purpose is to enhance classroom community social bonds, teach appropriate behavior, and give students opportunities to interact in their learning. These objectives bring into question how social skills develop and how they result in beneficial outcomes for students. The theoretical model of social skills embedded within a larger framework for the growth of adaptive behavior must also be integrated into a system of prevention for *RC* to be effective, drawing in established criteria for evidence based prevention.

The Definition of Prosocial Behavior

RC, and many other school-wide programs in the behavioral family, aims to increase a broad range of positive behavior in the classroom. Student classroom behavior can be defined as social-behavioral functioning; observable adaptive behavior that serves as an academic enabler (Malecki & Elliott, 2002) through the effective use of positive social skills and self-management strategies that facilitate social competency (Gresham et al., 2001) . This is different than social-emotional functioning, which is more internalized

to the child and focuses on attributions, feelings, and attention (Conduct Problems Prevention Research Group, 1999; Izard, Fine, Schultz, Mostow, Ackerman, & Youngstrom, 2001). While social-emotional competency is strongly correlated to student behavioral outcomes, it cannot be directly observed, making it difficult to evaluate as a criteria for evidence based prevention in schools

Caldarella and Merrell (1997) conducted a literature review of 19 studies with the purpose of providing a positively orientated, working definition of social skills that could be used to form outcome measures in treatment. Results indicated that social skills were measured across studies using one of three perspectives: skills were either determined by feedback by students within the environment of intended generalization (peer acceptance); previously established definitions generated by experts (behavioral definition); or direct observation of peer behavior that led to a reaction of positive, desirable, behaviors from peers (social validity). The social validity perspective was by far the most popular method across studies. An example of the social validity perspective would be a behavior that resulted in peers allowing a student to play a game the student wanted involvement in. A practitioner could observe this skill, break it down into teachable lessons, and instruct till mastery. The skill, when acquired, would also need to be taught for generalizability to facilitate performance.

Using a three level qualitative analysis, Caldarella and Merrell (1997) reduced social skills to five major dimensions: peer relationships, self-management, academic performance, compliance to authority, and assertion. Each of these dimensions operate on a continuum. The dimensions share characteristics of current assessment practices. For instance, the frequently used *Academic Competence Evaluation Scales* (ACES) by

DiPerna and Elliott (2000) contain questions that rate obedience, social skills, and academic performance. Ratings on these dimensions are used to gauge behaviors separate but necessary for effective instruction.

Social skills are only one component of a theorized bidirectional model of adaptive behavior. Grossman (1983) defines adaptive behavior as "the effectiveness or degree with which the individual meets the standards of personal independence and social responsibility" (p. 380). A significant deficit in adaptive behavior may be indicative of a disability, such as a developmental disability. In Merrell's (2008) model, social skills and peer relations interact to facilitate the development of social competence. Social competence is a critical skill that is a prerequisite of building effective adaptive behavior within a given context. This theoretical model is highly relevant to the mechanisms proposed to underlie *RC* and other universal social-behavioral programs. Such programs hypothetically facilitate learning through a higher quality and frequency of positive peer interactions which align with Merrell's (2008) model by improving antecedents of adaptive behavior.

Strong student social skills may also lead to more instructional time in the classroom. Time spent correcting antisocial behavior is reduced to allow more time for teaching. This is in addition to the prevention of mental health problems that are associated with social competency deficits. Merrell's adaptive behavior model also creates more salient variables to intervene on. Targeting the improvement of social skills and peer relations ultimately may address a student's adaptive behavior to the environment. A limitation of the model is its validity. Social skills have demonstrated notoriously hard to define as a broad construct since the skills involved are closely

linked to micro-level contextual factors and macro-level cultural norms that create, to some degree, unique performance variables in each situation (Bronfenbrenner & Morris, 2006).

Sheridan et al. (1999) offers a more concise definition that blends the contemporary understanding of social skills and social competency, "goal-directed, learned behaviors that allow one to interact and function effectively in a variety of social contexts" (p. 86). The definition emphasizes the use of discrete, observable behaviors and generalization of skills in intervention planning to specific contexts in which the social skill will be needed. In doing so, Sheridan et al. (1999) defined social skills with the express desire of directly linking social behaviors to assessment and intervention in a process similar to the problem solving model (Fuchs, Mock, Morgan, & Young, 2003) by addressing problems as environmentally based behavioral deficits. An assumption of such a model is that if contextualized deficits are not remediated early, the difference between the social performance of the student and social expectations will continue to increase in severity. Ford and Lerner (1992) argues in their developmental-systems theory that elapsed time between the emergence of a problem and treatment is proportional to the severity, depth and treatment resistance of the problem, further adding to the argument for social-behavioral prevention. Bronfenbrenner and Morris (2006), in their discussion of the Bioecological Model, would agree, citing the developmentally progressive complexity of learned behavior, leading to a rising challenge in mastering certain benchmark skills that is expected with maturation.

Context-relevance theory (CRT), introduced by Sailor, Goetz, Anderson, Hunt and Gee (1988), was originally introduced as a mechanism of increasing instructional

generalization amongst students with developmental disabilities, linking contemporary understanding of the regulatory function and instructional potential of the environment to school-based assessment and intervention. Sailor et al. (1988) viewed CRT as a technology that would eventually support and enhance the growing trend of mainstreaming. CRT has four major tenants. First, increasing the frequency of horizontal interactions by allowing students of similar performance levels and varying performance levels to interact and practice and model skills. Second, enhancing locus of control and motivation by allowing students opportunities to develop functional competence of skills as they are taught. Third, maximizing instruction by employing conditions of associated cues and effects where physical aspects of the learning environment are used as instructional cues. Finally, employing interrupted habitual chains of behavior. This final principal capitalizes on the target of intervention occurring as early in a behavioral chain as possible to enhance generalizability of later behaviors in any linear skill chain.

Influenced by CRT, Sheridan et al. (1999) discussed the need for a behavioral counterpart to curriculum-based measurement, where an assessment of the environment prior to an individual behavioral assessment is completed to align and contextualize behavioral objectives appropriately. CRT suggests interventions should be molded to the environments in which the target behaviors will be used. Social skills should be immediately relevant and reinforceable in a student's typical environment. To the greatest extent possible, the targeted social skills should be generalizable across environments. Additionally, contexts need to be tailored to allow ample opportunities to both practice and demonstrate mastery of a skill (Sailor et al., 1988; Sheridan et al., 1999). Sailor et al. (1988) noted that across several lines of research, varying levels of opportunity to

respond served as a stronger independent variable than varying reinforcer strengths. Prevention programs like *RC* may partially contribute to social-behavioral development through these mechanisms by embedding in-vivo instruction across all members of a given classroom community.

CRT applies readily to a school-based prevention framework and helps explain the effectiveness of high quality prevention. CRT would favor programs that operate within the environment of targeted change. School-Wide Positive Behavioral Support (SWPBS) is an example of a program that aligns with CRT. The first step in SWPBS is to carefully analyze the environment and develop context-specific goals and rules. Intervention is typically targeted at environmentally-based behaviors that are reinforced in vivo'. Tier II intervention efforts are done in collaboration with teachers so behavioral modification techniques are ubiquitous across environments in which the student operates.

Development of Social-Behavioral Programming Within a Responsive to Intervention Framework

A national reorganization of school assessment and referral practices has been spreading rapidly across the nation. The three-tiered Response-to-Intervention (RTI) model has moved from theory to federal legislation with the reauthorization of the Individuals with Disabilities in Education Act of 2004 (Reschly & Hosp, 2004; Zirkel & Krohn, 2008). Schools may now use RTI as a means to identify students with educational needs not met through the general curriculum, and as an alternative to more traditional test-and-place policies. RTI focuses on longitudinal data tracking, intervention based on

demonstrated student need within the curriculum, and constant revision and application of evidence-based prevention and intervention practices.

RTI emerged out of a growing frustration in the 1980s of a diagnostic system for identifying learning disabilities based on discrepancy analysis between potential ability and performance, termed the traditional model (Bradley, Hallahan, & Danielson, 2002; Fletcher, Fuchs, Lyon, & Barnes, 2006). The traditional model has been criticized for lacking treatment validity, delaying intervention, and compounding statistical error (Bray, Kehle, & Hintze, 1998; Fuchs & Fuchs, 2006; Reschly, 1988). Fuchs et al. (2003) noted that the introduction of the Education for All Handicapped Children's Act (PL 94-142) of 1975 created the impetus for schools to implement specialized services for students with disabilities, resulting in a need to segregate students with disabilities who were struggling from students without disabilities. The accepted solution to comply with federal law and facilitate diagnosis was a discrepancy based approach where specific idiographic quantitative scatter across abilities informed diagnostic decision making and treatment (Fuchs et al., 2003). While perhaps intuitive at the time, the method was fraught with limitations. Variations in state-adopted definitions of a disability created a sense of arbitrary judgment in evaluating discrepancies. Poor treatment validity of assessment tools led to inefficient intervention (Fuchs & Fuchs, 1998), and a surge of "mild" learning disabilities brought the system into question (Reschly, 1988).

Discrepancy based approaches also lacked sensitivity, often failing to pick upon on learning problems until several years into a child's education when the magnitude of the problem became considerable in comparison to the achievement of peers. When in 1988 Reschly called for a "school psychology revolution", the context was set for a more

accurate, efficient system of assessment and intervention. RTI emerged out of this context as a flexible, sensitive, strategy with a focus on instructional validity. RTI typically uses a problem-solving model, where assessment and consultation result in solutions unique to the presenting problem. The RTI model has flexibility in its use and can be used as two-level, three-level, or four-level system, and can include a wide range of assessment and diagnostic policies (Fuchs et al., 2003). In a three-level model, prevention-orientated, evidence-based instruction or general curriculum is given to all students.

Traditionally, RTI has been conceptualized within an academic instructional domain. Students could move up and down the continuum of intensity based on assessed need and response to previous interventions. Recent iterations of the model however have transposed the 3-tiered approach onto behavioral and social-emotional problem-solving (Gresham et al., 2001; Gresham & Project REACH, 2005; McIntosh, Chard, Boland, & Horner, 2006). Together, the two systems considered within a tiered structure simultaneously may be referred to as a dual-pyramid approach. Gresham et al. (2001) observed that the problems that plagued the assessment of learning disabilities and the corresponding interventions also have reduced the effectiveness of behavioral remediation in schools. Specifically, the reactive approach to remediating behavioral problems often occurs too late and at too low a dosage. Traditional behavioral assessment may lack treatment validity, resulting in a mismatch of problems to intervention. Another limitation of reactive behavioral assessment is the lack of generalizability of treatment; the distinction between teaching social skills in a contained environment and facilitating overall social competency that allow students to use skills in a fluid fashion in different

environments amongst different populations. This is commonly referred to as the "teach and hope" problem. While intervention may address a deficit in skills, the problem with performing in a natural setting has been largely ignored.

To improve on the model of service delivery for behaviorally based problems, the field has relied on lessons learned from academic remediation framed within RTI to develop a behaviorally orientated model (Gresham & Project Reach, 2005). More specifically, behavioral intervention can be organized into tiers of intensity. This suggests the equivalent of a core curriculum, a school-wide behavioral prevention program, must be implemented with fidelity. McIntosh et al. (2006) began to address a social-behavioral RTI with a descriptive study of the relationship between these two major domains of student development. The sample included a subset of students from six elementary schools, totaling 1,653 students from a district that was simultaneously implementing tiered service delivery for both behavioral problems and reading difficulties. SWPBS was implemented for behavior and a phonics based literacy program for reading. This including universal level screening measures that were used to progress monitor students in Tier II and Tier III interventions; Office Discipline Referrals (ODR's) for behavior and the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good & Kaminski, 2002) for reading.

Results of this study indicated that roughly the same proportion of students were distributed amongst the three tiers: 90% in tier I, 2% to 7% in Tier II, and 1% to 2% in Tier III. While overall behavior was more variable than reading, this suggests that the two domains potentially covary. In first grade a majority of students with more than 2 ODR's were also below benchmark for reading on the DIBELS. Limited by a lack of statistical

analysis, McIntosh et al. (2006) offers only potential hypotheses regarding the relationship between reading and behavior at various levels of service. The study was also only descriptive in nature, with no explicitly defined intervention or prevention efforts for either reading or behavior. Despite limitations, the attempt to understand how similar variation occurs across domains and interact helped forward the idea for a dual-pyramid model to student functioning.

Nelson et al. (2009) conducted a study along a similar theoretical line to McIntosh et al. (2006), evaluating only a behavioral tiered model. The study included hypothesis testing, the explicit implementation of behavioral interventions, and tracked students longitudinally, leading to more valid results. They also make the important distinction between a public health model, often used synonymously with RTI, and a behavioral model. A public health model, used in McIntosh et al. (2006), had students move through tiers of service to determine the level of appropriate intervention, starting with primary prevention of a problem and ending with tertiary intervention of a resistant problem. A behavioral model uses environmental or family based predictors of risk to match various intensities of prevention to the appropriate target sub-population instead of relying on resistance to treatment. The tiers range from universal prevention to prevent the expression of a problem to indicated prevention to reduce the severity and duration of problems already observed. In Nelson et al. (2009), a behavioral model was used with a multiple gating assessment strategy to appropriately place students by perceived risk for behavioral problems.

Four-hundred and seven students divided into four cohorts were followed longitudinally for up to two years. Students ranged from kindergarten through third

grade, drawn from seven elementary schools contained in a single district. Students were given universal screening using the first and second gates of the *Early Screening Project* (ESP; Walker, Severson, & Feil, 1995) and the *Systematic Screening for Behavior Disorders* (SSBD; Walker & Severson, 1990) and ranked based on score. The top five students who had the highest scores on the first gate were assessed using the second gate. Students who then scored below the 20th percentile on the second gate were enrolled in the selected (Tier II) prevention group. Students who were receiving services for behaviorally-based problems prior to the study were automatically enrolled in indicated level interventions. Interventions consisted of a cognitive-behavioral classroom management program for universal prevention, a teacher consultation and parent consultation program for the targeted sample, and multi-systemic therapy for students at the indicated level. Students were assessed at the beginning and end of the school year for two consecutive years using the *Social Skills Rating System* (SSRS; Gresham & Elliott, 1990) and the *Woodcock Reading Mastery Test-Revised* (WRMT-R; Woodcock, 1998).

Nelson et al. (2009) hypothesized that the behavior model implemented would keep SSRS and WRMT-R scores stable at the universal level using age-based norms, while students at the targeted and indicated level would show improvement over pretest scores, controlling for gender and SES status. Results were consistent with these hypotheses. All groups showed significant improvement across SSRS dimensions at post-test, and trend lines were relatively stable at a two year follow-up. Both positive and negative behaviors showed change in expected directions. However, in contrast to other research on the relationship between positive social skills and academic outcomes (cf.

Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Malecki & Elliott, 2002; Wentzel, 1993), the multi-level behavioral intervention had no significant effect on academic skills. In fact, age-based trend lines for academic outcomes were negatively sloped for the targeted and indicated group. The results brings into question the critical level of adaptive functioning required for academic performance to be affected, if at all.

Taken together, the results of Nelson et al. (2007) demonstrated the powerful effect a comprehensive prevention model can have on school-wide behavior; however, the study failed to replicate findings from past studies that modeled the significant relationship between behavior and academics. Additionally, the authors did not include a tiered system for academics like that of McInotosh et al. (2006). The results clarify the need for certain assumptions of a multi-tiered model to be fulfilled including (a) the use of evidence based practice at all levels and a clear understanding of the limits of effectiveness of any given intervention, (b) general and preventative efforts in this first tier that are implemented with fidelity and are scientifically based and supported, and (c) educational placement decisions suffer a lose of validity when inferences of student performance are made based on reactivity to an untested intervention (Fuchs & Fuchs, 2006).

It has become an expectation of school psychologists to understand intervention and prevention methods that facilitate the growth of social-behavioral competency (Ysseldyke et al., 2006). The promotion of social-behavioral skills has been linked to positive outcomes both at the student and classroom level. From a student perspective, the promotion of social skills has been shown to lead to higher reading and math achievement (Caprara et al., 2000; Malecki & Elliott, 2002; Wentzel, 1993) and less

frequent and less severe delinquent behavior (Gresham et al., 2001; Najaka, Gottfredson, & Wilson, 2001). From a classroom perspective, the promotion of positive behavioral skills creates a learning environment that maximizes instructional time, reduces time spent correcting behavior, and creates a positive, safe climate for student learning (Lewis, Sugai, & Colvin, 1998; Luiselli, Putnam, Handler, & Feinberg, 2005; Sugai et al., 2000).

Within the RTI literature there has been a substantial body of research addressing the psychometric properties of tier II and tier III interventions. However, there is a paucity of research at the tier I level that addresses the entire student body, particularly in the social-behavioral domain. Considering the substantial cost of such programs, allowing consumers to make wise choices addressing the specific needs of their population is critical. Furthermore, good prevention can serve as a wise investment strategy, saving schools thousands and the nation millions, on an individual student who's negative trajectory is detected and remediated early on (Biglan et al., 2004), whether it be a reading or behavioral deficits/excesses. As the federal government has endorsed RTI, it simultaneously has supported the development of criteria for empirically validated prevention strategies (Weisz, Sandler, Durlak, & Anton, 2005). For example, while specific studies of *RC* have fulfilled some of the standards for effective prevention, there are still constructs and standards left unexplored that must be investigated given the widespread dissemination of the program.

The Argument for Investment in Social-Behavioral Instruction

A recent focus within the contemporary literature on behavioral management strategies and the development of child social skills has shown how critical competencies

outside the academic curriculum are not only linked to crime prevention and psychosocial health, but also academic motivation, engagement, and achievement both in math and reading (Caprara et al., 2000; DiPerna, Volpe, & Elliott, 2001; Malecki & Elliott, 2002; Wentzel & Watkins, 2002; Wilson, Gottfredson, & Najaka, 2001). The importance of social skills instruction in the classroom can be traced back to the work of both Vygotsky (1978) and Bandura (1986) amongst other early researchers.

Bandura's social learning theory stipulates that learning can occur in the absence of direct reinforcement or consequence. Rather learning, in particular the acquisition and shaping of social behaviors, can be acquired observationally (Bandura, 1986). Rotter (1982) expanded upon social learning theory by discussing specific elements embedded in the sociocultural environment that influence the actuation of discrete behaviors. The complexity of social learning phenomena in a given classroom is vastly rich and complex. A student's behavior will be influenced by their own subjective psychological perceptions of the behavior of others, the expectancy of a reward based on the observation of others, and the culturally determined reinforcement value of that reward. Rotter summarized this relationship as a general equation where behavioral expectancy (*BP*) equals the function of the expectancy (*E*) of a reward (*R*) and the value (*V*) placed upon such reward ($BP = f(E \ \& \ RV)$). As a classroom's overall behavior deteriorates, the status quo' is altered and negative learning occurs. For example, if a student observes another student receiving attention or consoling for obstinate behavior such as tantruming; that student may then adopt that behavior, despite it previously not being in the student's behavioral repertoire, particularly if the behavior is culturally compatible. With a complex network of social learning occurring in a classroom it is critical that

prevention programs be put in place to guide overall group development in a positive direction. *RC* places heavy emphasis on the first six weeks of school to establish an early culture of positive behavior, falling in line with social learning theory.

Vygotsty (1978) emphasized the need for rich social dialogue in the classroom to facilitate child development. Vygotsky believed that the most effective learning occurs through a social discourse between peers or a student and teacher of various instructional levels in a process known as "scaffolding" (1978). Scaffolding is defined as the optimal match between the ability of a student and instructional level of a teacher or between peers that allows facilitation of learning that is neither too low nor too high for the student. Vygotsky's theory on the power of socially embedded instruction has been confirmed through contemporary research. For instance, peer-to-peer tutoring systems that use careful matching between abilities have been shown to be highly successful in increasing student understanding of content across academic areas (Mathes, 1994; Menesses & Gresham, 2010). For such a system to be effective, students need to be taught how to positively interact with their peers in a way that would be conducive to meaningful instructional dialogue, further necessitating the need for research of social-behavioral prevention programming.

Recent research has moved past theory on the benefits of strong social-behavioral competencies to confirmation through statistical modeling. Using time lagged structural equation modeling on a sample of 149 urban 4th grade students, Malecki and Elliott (2002) found that social skills - measured in the fall using the ACES (Diperna & Elliott, 2000) – significantly predicted spring level reading scores, measured on *the Iowa Test of Basic Skills* (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993). Teo, Carlson,

Mathieu, Egeland, and Sroufe (1996) found similar results in a longitudinal study spanning six years. They theorized that social skills represent a system of discrete skills that are learned and then built upon like academic learning. When intelligence was controlled for, social skills emerged as a significant predictor of achievement. Collectively, these studies point to a growing body of research that a student's level of social–behavioral functioning serves as a gateway to effective academic benefit. However this only further increases the need for programs to show acceptable levels of effectiveness before being implemented, particularly since the field is moving towards using a tiered model for social and behavioral problems. The potential for student growth must be confirmed beyond a doubt if instructional time is to be sacrificed to implement such programs.

Wentzel (1993) conducted one of the earliest studies on the relationship between social skills in the classroom and its relationship to other student outcomes in a cross-sectional study of 423 middle school students and their teachers using teacher-nomination to assess prosocial and antisocial behavior. Outcome measures included various demographics, the *Stanford Test of Basic Skills* (STBS; Harcourt, Brace, 1987) scores, student GPA's, and teacher rating of academic behavior using a short opinion based survey. Using regression with grade point average (GPA) and STBS scores as outcome variables, results indicated that both prosocial and antisocial rankings were significant predictors of GPA and that prosocial behavior was predictive of STBS scores (Wentzel, 1993). Both prosocial and antisocial behavior were also related to subjective perception of academic behavior. While results were promising, they were noticeably more subdued than the results of future studies measuring similar outcomes. This may be explained in

part by the use of longitudinal data and more reliable, normed, measures to narrow in on social skills introduced in later research.

In another study, Caprara et al. (2000) conducted a five year longitudinal study in Italy with 300 3rd grade general education students. Peer-nomination, self-nomination, and teacher-nomination were used to form latent variables of prosocial and aggressive behavior. A subsample of students were assessed for academic achievement using a standardized country-wide achievement battery, which were than entered into a latent growth model to predict academic achievement and social preference in 8th grade. Surprisingly, the authors found that the only significant predictor of student achievement in 8th grade was ratings of prosocial behavior in 3rd grade (r = .57). Prosocial behavior also was the most powerful predictor of popularity at the five year follow-up. While the subsample of students who had academic achievement measured was small (n = 100) in comparison to the overall sample and as a result, possibly underpowered the study, the powerful effect of prosocial behavior as a mediating variable for academic achievement merits attention by practitioners.

Taken together, the results of Caprara et al. (2000), Wentzel (1993), and Elliott and colleagues demonstrates converging evidence that the focus of intervention should be primarily on facilitating positive behaviors, not the reduction of negative behaviors. One of Skinner's earliest observations on operant conditioning was that using reinforcement to increase the frequency of positive behavior was far more effective and durable than using punishment to suppress negative behavior (Kendler, 1987). This hypothesis was confirmed in a federally funded meta-analysis by Carr et al. (1999) and is the guiding principal behind the popular field of positive behavioral support (PBS). *RC* texts devote

several chapters to the discussion of positive teacher language. Teacher language in these texts is taught to be positive, behavior-specific, and make use of lots of encouragement. *RC* however does not condone the use of tangible reinforcers, instead encouraging teachers to foster children's internal motivation. This stands against some of the basic tenants of PBS (Carr et al., 1999). PBS would encourage the use of both verbal praise and tangible rewards to create a comprehensive reinforcement plan potent enough to change behavior based on the needs of the student.

Malecki and Elliott (2002) conducted a similar longitudinal study across one school year to that of Caprara et al. (2000). They hypothesized that positive social skills and problem behavior represent two sides of a latent dimension they called "Academic Competence" which was hypothesized to have a significant relationship with student academic achievement over time. They measured academic competence with the SSRS and academic achievement using the ITBS on 139 3rd and 4th grade students. Both the SSRS teacher and student versions were used. Results indicated a moderate correlation of fall teacher-rated prosocial skills to spring achievement. Within the fall dataset, variance in ITBS scores were not significantly accounted for by problem behavior, similar to the findings of Capara et al. (2000) and Wentzel (1993) that only level of positive behavior is correlated with achievement. Interestingly, teacher rating of social skills and problem behavior significantly outpredicted students self-assessment of these skills in relation to academic achievement. This finding was replicated by DiPerna, Volpe, and Elliott (2005) in a sample of 394 students across elementary grades in the Northeast United States. In a path analysis using the ACES as a dependent variable, it was found

interpersonal skills was significantly correlated to reading scores, mediated by academic motivation and engagement.

The importance of prosocial behavior has been shown to expand beyond academic achievement and academic enablers. A meta-analysis by Najaka et al. (2001) highlights the importance of prosocial development in preventing delinquency. The meta-analysis included 87 studies that evaluated universal prevention programs to reduce problem behavior in schools. Targeted behaviors included academic performance, bonding to schools and social skills, which were then regressed onto the outcome variable of problem behavior. Bonding to schools was by far the most powerful predictor (r = .84), followed by social skills (r = .11). However, the authors note that a confound emerged with the measure used to quantify social skills. Studies that relied on self-report showed highly variable, non-significant results (r = .03), while studies that used peer report or teacher report had a strong negative correlation with problem behavior (r = -.60). This insight demonstrates that observed social competency is linked to individual higher achievement, increased popularity, reduced delinquency in schools, and better academic competency and is most accurately observed by the respective classroom teachers.

Prosocial behavior is also negatively correlated with school-level outcome variables such as school-wide behavioral infractions, problem behavior, and suspensions. Lewis et al. (1998) conducted a multiple baseline single-case study on transition environments of an elementary school. The researchers consulted with teachers to introduce a targeted social skills instruction program and operant conditioning strategies utilizing positive reinforcement over the course of one school year. Here, operant

conditioning could be considered as a type of social skills intervention since the consistent application of conditioning shapes behavior that may include social skills; in this case increasing desirable behavior including prosocial behavior. In a follow-up analysis, Solomon, Klein, Hintze, Cressey, and Peller (2010) converted the results of this study to proportion of variance effect sizes and found a strong effect size of $r = .69$ for reduction of problem behavior, measured using direct observation. The reduction of problem behavior may lead to a reduction in transition time, more exposure to the curriculum via reduced principal visits and reprimand time, and a reduction in school resources devoted to handling problem behavior.

In summary, research on child social-behavioral functioning has led to a theoretical model for the development of positive behavior in context and in consideration of social skills' role in adaptive functioning, academic outcomes, and delinquent behavior. Overt behavior that is advantageous to learning and peer relationships can be considered within a contextually based causal model for development. Social skills are but one necessary component of social competency grounded in a social validity perspective; the effective use of overt social skills that leads to positive peer responses. Social competency ultimately may lead to adaptive functioning to one's environment (Merrell, 2008). This positive adaption has been shown to be a critical enabler for academic instruction, and in least two empirically rigorous studies, predicted achievement in later grades better than early grade achievement (cf. Caprara et al., 2000; Najaka et al., 2001). Early theorists such as Vygotsky (1978) and Bandura (1982) predicted the role of social skills in facilitation of instruction and the mediating role of community based contextual variables in this process.

While this fairly specific model for adaptive behavior has demonstrated evidence, how it can be implemented into a system of prevention and intervention remains another question entirely. The adaptive behavior model needs to be considered within the constraints of limited school resources and limited flexibility. Additionally, the relationship between behavioral and social dimensions may not remain equal at all levels of intervention (i.e., universal, targeted and indicated levels) and may vary based on demographic factors, necessitating the need for further research considering both behavioral and instructional models.

Standards of Practice

The RTI model requires that specific assumptions have been met in regard to effectiveness and fidelity of the prevention and intervention efforts used (Fuchs & Fuchs, 2006; Fuchs et al. 2003). As RTI increases in popularity, it becomes ever more important that curriculum and intervention technology match pace, and their generalizability gauged before widespread implementation. However, such research on program effectiveness has lagged behind the implementation of RTI, particularly at the Tier I level of core general education programming. In an RTI model, Tier I should make use of programming and universal prevention at the school-wide level to bolster student resilience when a low dosage of service can be effective on a more malleable early developmental level. This includes the core curriculum, school-wide behavioral policies, and embedded social skills instruction (Gresham et al., 2001; Sprague & Walker, 2005; Wilson et al., 2001). To answer such questions, not only does an expanded research effort

need to address Tier I program effectiveness, but also the advancement of appropriate tools to measure fidelity and effectiveness of Tier I programs in the classroom.

The *No Child Left Behind Act*, Title IV, addresses the benchmarks required for effective social-behavioral prevention: the need for robust evidence prior to implementation, constant parent involvement, and use of formative assessment measures to gauge effectiveness (Sprague & Walker, 2005). Meeting these benchmarks is required for a school to invest public federal money in a program. Hence there is a substantial legal weight placed on establishing criteria for evidence-based social skills programs. This aligns with prevention fundamentals discussed in the field previously, such as that of Sprague and Walker (2005) who outline five necessary components of behavioral prevention: systematic direct instruction, academic restructuring of the environment, positively based interventions, early screening, and alternatives to severe negative consequences such as expulsion.

Sprague and Walker (2005) discuss the need to create an environment that favors student social-behavioral gain. Along with aligning with principles of effective prevention, including sensitivity to socio-culture need, overlapping support, and at minimum a partial script (as opposed to a program built purely on foundations and beliefs), the authors discuss the need for schools to foster an educational environment that favors the unique demands of effective social development. Empirical research to date has shown social-behavioral prevention needs to be provided in vivo', in contrast to pull-out services. Gresham et al. (2001), summarizing six narratives and six meta-analytic reviews on social skills training (SST), found that social skills training (SST) provided highly variable effect sizes, ranging from d = .20 to d = .87, with an average effect of d =

.48. These effects were heavily moderated, with SST for withdrawn students appearing more effective than SST for aggressive students.

Several consistent findings were noted in terms of the general limitations of SST. For one, it typically occurred far too late in child development; the average age of intervention was 12. Gresham et al. (2001) state that interventions after the age of eight require far more intensity than is typically, or in some cases possibly could be, delivered to be effective in a school environment. Referring back to the RTI model, academic difficulties also share this characteristic of becoming increasingly more difficult to remediate as the student gets older and the deficit gap increases, commonly referred to as "The Matthew Effect" (Stanovich, 1986), similar in operation to Ford and Lerner's previously mentioned Developmental Systems Theory (1992). Gresham et al. (2001) report that SST typically was isolated from natural contexts and while it may facilitate development of social skills, was not effective at teaching students how to use specific social skills at the appropriate times, an issue of generalizability and social competency. Furthermore, few studies reported on the fidelity of intervention implementation. Gottfredson et al. (2000) also noted descriptively that in their meta-analytic review, only 41% of were exposed to an implemented universal prevention program, with a tendency for a plethora of simultaneous, poorly implemented programs as opposed to the implementation of one, comprehensive, evidence-based program with high fidelity and maximum coverage of the target population.

Similar in scope to the Matthew Effect, the "Kindling" Hypothesis also supports the frontloading of prevention efforts, academic or behavioral, as more effective than a reactive treatment system alone. The Kindling Hypothesis has its theoretical origins in

stress sensitization models conducted in animal laboratory studies (Monroe & Harkness, 2005). In these early studies, it was found that the critical magnitude of a stressor, such as a mild electric shock, which triggers a cascade of neurochemical reactions of both short and long term changes, can change based on past events. The result is that successive trials of progressively lower levels of a stressor can trigger the neurochemical reaction that initially only a large shock could. Translated and then validated as a behavioral model, the Kindling Hypothesis states that once a major life stressor has triggered a depressive episode, subsequent episodes can be triggered by milder negative stressors, perhaps to the point where the anticipation of stressors or sub-detectable minor stressors can trigger full depressive episodes (Post, 1992). Essentially, the conditional probability of a major life event in triggering depression based on prior depression history is reduced; it is either subsumed by another function that causes depression or the effect of a major life stressor is diluted by more frequent, milder stressors that trigger depressive episodes (Monroe & Harkness, 2005).

The implications for the Kindling Hypothesis in regards to school-based student functioning is significant. If a student where to experience a major life event at school, such as a social trauma or academic failure, the child would become increasingly susceptible to internalizing episodes over time. Eventually, the treatment needed to not only end an episode but reverse, if even possible, the sensitization to stressors would be intensive therapy proportional to the length of time the issue has existed. Based on Monroe and Harkness' (2005) literature review of psychopathological development predicted by the Kindling Hypothesis, it is likely the severity of the sensitization would be far beyond the remedial capabilities of a school-based staff. On the other hand, if the

initial depressive episode was prevented, perhaps through effective pro-social strategies embedded within classrooms, the student's resilience to stressors of any intensity potentially remains unchanged, if not strengthened. The implementation of a tiered model of service delivery would allow quick response to minimize sensitization.

Gresham et al.'s (2001) review suggests that providing social skills training in the classroom, a natural environment for their use, and providing prevention efforts early on, are cost-effective solutions to the deficits in current remediation. This is further reinforced by lessons learned from empirically validated theories of acquired disability such as the Matthews effect or the Kindling Hypothesis. *RC* appears to align with some fundamentals of effective universal prevention programming, such as in vivo' practice and early intervention of difficult problems (it can be implemented as early as kindergarten). However, *RC* also stands in contrast to other fundamentals, such as having a script to guide lessons as opposed to a system of beliefs and having specific, measurable, objectives. On the contrary, *RC* encourages a unique, creative solution for each student with little emphasis on formative measurement and treatment integrity.

The Link to Good Instruction

It was mentioned previously that social-behavioral development is theorized as a gateway for academic success (McIntosh et al., 2006; Sugai, Horner, & Gresham, 2002). This hypothesis is reasonable under several different lines of reasoning. For one, students who have low social-behavioral functioning or difficulties that cannot be effectively managed by the teacher may be subject to punitive teacher responses, such as frequent time-outs or an ODR. This reduces the student's exposure to the curriculum, thereby

reducing achievement for the student with behavioral challenges. Additionally, within-class the student may express behavior that blocks their access to the curriculum, such as frequent externalizing behaviors including tantruming and hyperactivity, or internalizing behaviors such as anxiety or inattention.

Classroom social norms tend to vary on a group-level. As a result of the visible behavior of one particular student, or several, teachers may have to adapt their instructional approach, most likely to the detriment of the curriculum. This can include reducing instructional time to provide frequent environmental and behavioral management, such as behavior correction (e.g., “Jimmy, I will not tell you again to sit down”, “Jane, please stop horsing around with Bobby and get back to your work” or “Class! This is the last time I will ask! Everyone get to their seats now!”). Each of these transgressions is a moment lost out of delivery of the necessary instruction needed for student success. For instance, if a teacher spends fifteen minutes of each school day correcting negative behavior, this accumulates to over six school days of lost instructional time over the course of the year.

Poor student behavior also limits the quality of peer to peer interactive learning and usable teaching strategies. If students cannot interact effectively on task, the teacher may be constrained to a restrictive spectrum of teaching methods. The use of social-behavioral prevention in this context can be seen as an instructional investment. Early prevention efforts that take away from instructional time may reward an exponential amount of instructional time later in the year by reducing problem behavior.

While *RC* emphasizes social-behavioral growth within students, it is intended to address multiple levels of student functioning through a change in teacher behavior. If

such is the case, direct observation of a change in teacher behavior would be instrumental to validating the overall effectiveness of the *RC* program, and useful for other programs within the social-behavioral family.

The importance of teaching behavior emerged in the peer-reviewed literature as early as 1963 when Carroll published his theoretical model of critical variables that effect school learning. Carroll identified several types of time allocations in the classroom: the time needed to learn, the time allocated to learn, and actual time engaged in learning. These learning times in addition to individual student aptitude, perseverance, and ability to understand instruction, constituted Carroll's pioneering understanding of contextual variables that moderate student learning. Carroll focused on the time spent learning as a variable more easily manipulated than student aptitude, stating that ""aptitude" is regarded as relatively resistant to change, whereas it is the hope of the psychologist that he can readily intervene to modify "perseverance", "quality of instruction," or "opportunity for learning" (p. 731).

This understanding of learning time as a variable that can be altered to improve student outcomes at a classroom scale influenced applied teacher consultation. An example is the Beginning Teacher Evaluation Scale (BTES; Fisher et al., 1978), which used Academic Learning Time (ALT) as a primary outcome measure of effectiveness of new teachers. Repeated studies of ALT in the 1970's showed that time spent on instruction can be surprisingly low for some classrooms; Fisher et al. (1978) reports as low as 38% of classroom time. It is suggested that ALT be used as a formative tool in teacher consultation and that data be gathered through teacher maintained logs. Relevant to the current study, Fisher et al. (1978) also suggested that teachers use well planned,

explicit instructional lessons to minimize time spent correcting behavior, thereby increasing overall instructional time.

Contemporary empirical research has revisited the concept of instructional time as a potential formative tool for consultation. For example Vannest and Parker (2010) have developed a web-based survey completed by the teacher after every hour of instruction. The program codes teaching behavior in one of ten ways: academic instruction, nonacademic instruction, instructional support, responsive behavior management, preventative behavior management, special education assessment, state-mandated assessment, classroom assessment, special education paperwork and general education paperwork. These authors piloted the program on 31 special education teachers over nine weeks of instruction in the fall. Teachers were given access to the program and encouraged to complete the survey as many days as they could.

The authors then conducted a trend analysis to test the reliability of the instrument and the amount of days needed to develop reliable scores. Academic instruction was the most commonly coded behavior, followed by different types of paperwork. However, the use of time was highly variable teacher to teacher; the main effect for teacher far outweighed the main effect for time across behavioral categories. Furthermore, it was discovered that there was little variance in instructional time over time; behavior largely remained constant across teachers. Furthermore, change in behavior ranging from 44% to 245% was necessary to demonstrate significant change in behavior. The study demonstrated that instructional time can be reliably and quantitatively recorded and can be used for formative assessment. Furthermore, teaching time is not a constant across teachers; there are a significant population of teachers who may need consultation

regarding their use of instructional time. Unfortunately, this study focused on special education, not regular education. Additionally, the instrument was a survey completed by the teacher, not direct observation by an independent observer. Direct observation, hypothetically completed by an objective observer ad actively encoded, may be inherently more reliable than an indirect measure of behavior. This finding is supported by extant research. For example, Briesch, Chafouleas, & Riley-Tillman (2010) reported that systematic direct observation of students was more reliable than single-item behavior rating scales completed daily on students in a study utilizing both methods concurrently across 12 kindergarten students over ten school days. Twenty percent of the variance in rating scales was accounted for not by the true variance of the student, but rather by the individual characteristics of the rater. Finally, instructional time in Vannest and Parker (2010) was not correlated with the academic achievement of students.

Carroll's Model of School Learning and the BTES studies illuminated the potential of instructional time as an outcome measure in education. Despite these early studies, instructional time rarely has been quantified and used an outcome indicator in the contemporary literature. In particular, few studies have partitioned classroom time based on the behavior of the teacher. Only recently has the amount of teacher-led instructional time been used as a formative assessment of teacher behavior (Gibson & Hasbrouck, 2007; Vannest & Parker, 2010). Understanding how instructional time is used is critical to understanding the efficacy to effectiveness potential of *RC*, which is largely dependent on teachers controlled use of effective instructional time and environmental management at different points in the year.

Current Research on RC

RC has been evaluated by several different researchers using a variety of methods. However, the research is limited in terms of its generalizabilty to schools outside the controlled experimental conditions. Additionally, while observations of teacher behavior have been done in the past to assess fidelity, there has been no analysis of how teachers operate in the classroom based on a more universal scale, comparable to other prevention programs, including research on temporal shifts in teacher behavior across the school year.

A program evaluation done by Elliott (1999) showed modest implementation fidelity (23% - 88% depending on specific *RC* component), modest academic gains, and moderate social skill gain using the SSRS (teacher, parent and student versions were used in this study). The SSRS measures social skills, problem behavior, and academic competence. Social skills on the SSRS are further broken down into various subscales: Cooperation, Assertion, Responsibility, Empathy, and Self-Control. Elliott used a mixed design, tracking students both longitudinally and across schools for two years of program implementation. The experimental school was described as an urban Title 1 school of mixed ethnicity located in the Northeast, serving over 400 students in grades one through five. The demographics of the control school are not described. A sample of 300 students was drawn from across both locations, however only 66 students were tracked longitudinally for the full study duration of two years. Additionally, 34 teachers and 102 parents participated.

Teachers were given the SSRS-teacher, a *RC* fidelity survey, the ACES, and the *Student Self-Concept Scale* (Gresham, Elliott, & Evans-Fernandez, 1993). Parents were

given the SSRS-parent and students were given the SSRS-student and the ITBS. Outcomes measures were d = .41 for the SSRS-teacher, d = .07 for the SSRS-parent, and d = .34 for the SSRS-student over the course of one year of the study. Using data presented in this source, an effect size was calculated for year one differences for the ITBS, which was moderate in size (d = .31). Elliott evaluated ITBS results using a MANOVA, finding significant differences across all academic domains in favor of the experimental condition (p = .0001). No F-value was reported nor significance testing for other variables. This lack of significance testing for the ACES, the *RC* fidelity measure, and *the Student Self-Concept Scale* limit inferences that can be drawn from this study. Additionally, sample sizes were very unbalanced (n_e = 113, n_c = 34) for some dependent variables, which may have inflated the chance of Type I error.

A similar quasi-experimental, mixed design study by Rimm-Kaufman et al. (2007) showed moderate effect sizes in reading and math. Rimm-Kaufman et al. (2007) extended the sample population to three experimental schools and three control schools within a single district with an overall experimental sample of 759 students and 43 teachers in grades one through four and a control sample of 769 students. The student sample was overwhelmingly Caucasian in ethnicity, despite substantially more diversity reported in the district population. The experimental group's teachers were given both the *RC* level one training during the summer and the *RC* level two training during the following school year. Outcome variables were the *Connecticut Mastery Test-Math* (CMT-Math; Connecticut State Department of Education, 2006) and the *Degrees of Reading Power* test (DRP; Touchstone Applied Science Associates, 2002). The *CMT-Math Test* is described as a standardized assessment used for statewide testing in

Connecticut. The DRP is a nationally normed comprehension test for elementary students that uses a missing word format. Groups were roughly equivalent at onset of the study. The authors did not detail exactly when and how teachers were trained in *RC* (fidelity checks were later done), creating a concern for low content validity, exacerbated by the lack of random selection. Additionally, the author's used ANCOVA across groups, which may have increased Type I error by having a convenience sample with the pretest as control.

The authors used multiple ANCOVA's across groups for each year the study was in place. Findings from ANCOVA's were converted to effect sizes: $d = .16$ for reading and $d = .39$ for math when baseline was compared to the three year post-test. Interestingly, it took two full years of implementation before significant effects were observed. A between-groups fidelity survey showed significant differences in use of *RC* teaching behaviors between conditions ($t = 5.22$, $p < .001$).

In another study by Rimm-Kaufman and Chiu (2007) using a subset of the sample used in Rimm-Kaufman et al. (2007), two year differences between groups using different instruments than the original study were compared. The *Mock Report Card* (Pierce, Hamm, & Vandell, 1999) was used to test overall student achievement. It asks teachers to rate students as if they were being typically graded on a variety of reading and math areas. The *Student-Teacher Relationship Scale* (STRS; Pianta, 1992), a normed, Likert- style questionnaire for teachers, was used to evaluate the strength of the personal relationship between teachers and their classroom students. The SSRS was used as was the *Social Competence and Adjustment Scale* (Ladd, Profilet, & Muth, 1996), a Likert-style questionnaire that asks teachers to rate student classroom behavior in comparison to

peers. The authors' used hierarchal regression, deciding against controlling for nesting of students due to the highly variable n of each cluster in classrooms, despite the significant effect this nesting may have had on results. Reported effects of *RC* were largely non-significant. When pretest scores and family risk status were controlled, r^2 effect sizes ranged from 0 to .06 for the effect of *RC* on outcome variables. The strongest relationship was between teacher closeness as measured on the STRS and reported use of *RC* practices.

Combined results of these studies lack consistency and do not address key components of evidence-based practice that would allow early efficacy trials to move to such rapid dissemination (Flay et al., 2005; Nation et al., 2003). Standards for evidence-based prevention have begun to emerge and just like any other psycho-educational investment, prevention expenditures can be wasted if not wisely spent based on the merits of the program, how a program matches the population's needs and the substantiated transition from efficacy trials to effectiveness (Flay, 2005; Nation et al., 2003).

Nation et al. (2003) conducted a literature review of prevention program qualities in the mental health field that was generated out of a collective effort of the APA task force on prevention. The authors combined the results of 35 review articles that outlined key criteria for establishing effectiveness for prevention programs, generating 252 criteria. This list was then rank ordered using expert analysis, which resulted in nine criteria organized into three broad dimensions: program characteristics, matching the programs to target population, and implementation and evaluation of the program. Notable criteria consistent across studies and ranked as critical by experts included the use of a multi-method, multi- instructional approach; an informed, appropriate dosage

level; theory driven practices; a focus on developing positive skills instead of, or in combination with, reduction of undesired behavior; appropriately timed and culturally relevant practices; implementation of an experimentally sound outcome evaluation and thorough training of staff. The authors also noted a paucity in universal prevention research in contrast to prevention efforts directed at identified at-risk populations.

Current research on *RC* fulfills few of these criteria. Research has not established an effective dosage levels, has not matched results against possible demographic moderators, does not adequately assess fidelity, and is limited in social scope by the lack of generalizability studies. Program evaluations were done under the auspice of the developers of the *RC* program and sampled from pilot schools in close proximity to the NEFC. Work by Rimm-Kaufman and colleagues also was conducted in direct connection to the NEFC. An effectiveness study that measures the effect of *RC* in an "everyday" teacher population has yet to be completed.

Important questions left unanswered from past studies on the effectiveness of *RC* include: how does the heavy investment in classroom organization and social-behavioral competency relate to end of year student behavior and academic achievement? Will a new study show replication of results across academic and social domains when controlling for the nesting effects of students within classrooms, using different instruments and novel environments? Finally, how does *RC* change teacher behavior?

Proposal of Study – Rational and Purpose

Taken together, extant research suggests that investment in students' positive social-behavioral growth at an early age is critical to a variety of student outcomes,

including academic achievement. Framed within a prevention based tiered service delivery model, responsive and appropriately leveled intervention minimizes the magnitude of behavioral deficits and maximizes instructional time. Behavior, like reading, often does not develop along a positive trajectory by due course alone, validated by the prevalence of school-based behavioral problems in recent national samples (Hoagwood & Erwin, 1997) and the rapid rise of adolescent antisocial behavior in recent years (Satcher, 2001). To this end, desired behaviors must be broken down into their observed, socially validated components, and taught explicitly in lieu of, or in addition to, the suppression of negative behaviors.

At the same time, one cannot assume that prevention is necessarily preventative. The relationship between prosocial behavior, delinquency and academics must align with what is known about effective environmental-behavioral change. Specifically, prevention efforts must fulfill the criteria for evidence-based practice, most notably defined by the APA Taskforce on Prevention (Nation et al., 2003) and the Society for Prevention Research (Flay et al., 2005). Seemingly contrary to the recommendations of these organizations, many universal level programs have failed to reach criteria for dissemination despite their widespread use (Gottfredson et al., 2000). Payne et al. (2006) modeled these criteria for effectiveness using a sample of 504 schools, incorporating administer opinion of outcome effects. From this sample, principals and prevention coordinators were surveyed regarding the fidelity, coordination, and implementation of school-based prevention programming in their schools. Using structural equation modeling, the authors created a latent model that overlapped significantly with standards of quality mentioned previously and was statistically robust (*NNFI* = .88, *CFI* = .86, χ^2 =

695, *df* = 278). The duration and frequency of prevention sessions were shown to be highly correlated with implementation quality (β = .33) as was various demographic qualities of the population (β = .22), confirming the need to be mindful of potential social and cultural moderators (Payne et al., 2006).

RC has shown adequate efficacy but lacks demonstrated effectiveness. This is a particular concern given the high cost of training. Drawing from the criteria from Nation et al. (2003), *RC* has not been shown to be socio-culturally relevant in a wide variety of circumstances, has not shown to be of an intuitive sufficient dosage, and may not involve well trained staff outside of highly controlled studies. Drawing from Flay et al. (2005), *RC* has not been shown to operate effectively in "real world" conditions, generalizability has not been evaluated, and critical level for dosage response has not assessed.

Ultimately *RC* works to change teacher behavior to better facilitate the development of prosocial behavior and classroom management, which is hypothesized to then prevent mental health problems, aggression, and increase academic achievement through processes mentioned above. This study proposes to add to the literature on *RC* by both replicating and extending past research by examining whether exposure to the *RC* program changes the instructional practices of teachers, whether these instructional practices vary, leading to different levels of effectiveness, all while controlling for the potentially nesting effects of classrooms.

Research Questions and Hypothesis

Based on previous literature and the rational for this study, the following questions are to be addressed:

- Will results of an effectiveness study converge with previous findings on *RC*?
- *RC* claims to change teacher behavior, particularly in the beginning of the year, to better facilitate communication, routine, and positive community based skills in students through the mechanisms of adaptive behavior. By addressing these skills early in the year, theoretically academic enablers are developed before instruction takes place to maximize instruction later in the year both between the teacher and students and between peers. How does initial investment in classroom organization effect end of year instructional practices?
- Finally, since *RC* uses a consultee-based training model, changes in student behavior should equate to a change in teacher behavior. Can the observation of instructional practices be used to predict student level outcomes in the spring across a continuum of *RC* fidelity?

In regards to these questions, the following hypotheses are proposed for this study. Teachers who use the *RC* method, measured as a continuous variable using the Classroom Practice Measure and a *RC* fidelity survey used in past research (Rimm-Kaufman & Sawyer, 2004; Rimm-Kaufman et al., 2007; Rimm-Kaufman & Chiu, 2007), will spend more time than teachers who do not use *RC* investing in classroom structure and classroom community in the fall. To this end, teachers who spend more time in the fall investing in classroom community and classroom structure will show a greater ratio of time spent teaching to behavioral corrections in the spring. Both of these hypotheses will be tested using the Teaching Observation Tool (TOT), a momentary time-sampling observation of teachers in their classroom (Marcotte et al., 2010). Finally, teachers who use *RC* will rate student behavior using the ACES as improving at a steeper slope over

the course of one school year than teachers that don't use *RC*, or do so minimally. Specifically, the areas of math and student interpersonal skills will show the most growth, as suggested by previous research (e.g., Elliott, 1999; Rimm-Kaufman, Fan, Chiu, & You, (2007).

CHAPTER II

METHODS

General Method

This study used hierarchical linear modeling with two levels (teacher group, individual student difference scores). The effects of *RC* were measured over time while simultaneously controlling for the effects of having sample students nested within classrooms. Such multi-level modeling allows not only analysis of student level change, but also the primary target of change for *RC*, the teacher and change in teacher's behavior in the classroom. In regard to student level outcome variables, the current study used four sub-domains of the ACES: motivation to learn, reading achievement, math achievement, and social skills (DiPerna et al., 2001; DiPerna et al., 2005; Elliott et al., 2004). To measure the shift in teaching behavior from environmental control to instruction, the *Teaching Observation Tool* (TOT) was used (Marcotte et al., 2010). Both these measures were completed twice, with the primary variable of interest being the change from fall to spring. Additionally, demographic information on teachers was gathered. To quantify fidelity of *RC* implementation, a modified version of the *Classroom Practice Measure* (CPM) was employed (Rimm-Kaufman et al., 2007).

Sample

Twenty-four teachers and 178 students participated in this study. Teachers were recruited from nine different elementary schools. Teachers ranged from having no exposure to *RC* to having completed a weeklong workshop. Ten teachers completed a

daylong workshop (which includes reading materials) as their highest level of *RC* training. Twelve teachers had attended a weeklong workshop (also including reading materials). Two teachers had either no *RC* exposure or had familiarized themselves with an *RC* textbook only. The majority of teachers had a Masters in Education (n = 16), with the remaining having Bachelors level training (*n* = 5) or a terminal degree beyond a Masters (*n* = 1). Teachers had an average of 11.21 years of teaching experience (*SD* = .701). A distribution of included grade levels is presented in Table 1. There were 97 female students and 81 male students. Class sizes ranged from 8 to 22 students, with an average of 15.70 students (*SD* = 3.60).

Table 1

Cumulative Frequency of Grade Levels

Grade	Frequency	Percent	Cumulative Percent
Kindergarten	4	16.7	16.7
1st grade	6	25.0	41.7
2nd grade	6	25.0	66.7
3rd grade	2	8.3	75.0
4th grade	3	12.5	87.5
5th grade	3	12.5	100.0

Six of the schools were located in a district located in Western Massachusetts. The district served a total of 6,072 students taught by 508 teachers. Forty-nine and seven-tenths percent of the student body was considered low-income. The district was majority

Caucasian (76.3%). Standardized test scores for the district are descriptively identified as "high" for language arts and "moderate" for mathematics. In comparison to state scores, the district ranks slightly below average, although it met AYP in the last academic year (2008 – 2009). Seventy-one students and ten teachers came from this site.

A second location was a single public elementary school located in Western Massachusetts, in a separate district. The school served a total of 410 students, 17.1% of which were considered low income. The district was primarily Caucasian (84.1%). The school was considered to have "high" performance in both language arts and mathematics on statewide assessment. Annual Yearly Progress (AYP) was met for language arts in the previous academic year, but not for mathematics. Forty-one students and six teachers came from this school.

A third location was a private school located in Eastern Massachusetts. The school served a total of 270 students from Kindergarten through 8th grade. Twenty-four percent of students received financial aid. The school reported that 85% of students were Caucasian. Fifty-six students and seven teachers were recruited from this school. Standardized test scores were unavailable at the time this study was conducted.

The final site for this study was an urban charter school located in Providence, Rhode Island. The school has 246 students. The majority of students were Hispanic (43%), with a sizable minority African-American (31%), and then Caucasian (17%). The school was below state averages for reading and writing; 18 percentile points and 22 percentile points respectively. The school did make AYP in 2008. Sixty percent of the student body was considered low-income when free and reduced lunch status were used

as a proxy. One teacher and eight students came from this site (a second teacher from this site dropped out of the study mid-year due to maternity leave).

Measures

Teaching Observation Tool

What is "good instruction" and is it possible to create an instrument that can measure teacher behavior to answer this question? This study used a 30-minute, momentary time-sampling observation tool, the TOT, that is hypothesized to be sensitive to the instructional practices of the teacher (Marcotte et al., 2010; see Appendix A). The TOT is based on the work of Gibson and Hasbrouck (2007), who developed a brief observation using a frequency count that categorized teaching behavior as either managing the classroom environment, delivering instruction, or correcting behavior. The observation was intended to measure change in teaching behavior over time during classroom consultation, with the ultimate goal of shifting the most time to small group instruction.

Similar measures have been used in past research and have demonstrated that teaching behavior can be quantified and does change over time. For example, Connor, Morrison, and Katch (2004) found that, using a descriptive measure of teaching behavior, teachers use of student-directed or teacher-directed instruction and use of explicit or implicit instruction at the beginning of the year affected end of year teaching practices. Additionally, there was an interaction between shifts in teaching behavior and the entry-level ability of the students on end of year student achievement. Students with low entry-level achievement responded better to explicit, teacher directed instruction than students

with high entry-level ability. The study demonstrated that teaching behavior can be quantified and change in teaching behavior is potentially significant over time. However, the use of a descriptive measure, as opposed to systematic direct observation, raises concerns for reliability of the independent variable in this study.

The TOT has a 15 second interval with a three second observation time. During that time, the observer codes teacher as either: "teaching" in small or whole group; "feedback", defined as giving academically-orientated feedback directly to a single student or group of students; "environment", which is defined as managing the classroom such as directing students to gather supplies or line up at the door, or "behavior", which is defined as either action or verbal behavior directed to correct a student or group of students who are not performing to teacher expectations or verbal or non-verbal (i.e., marking a sticker chart) recognition of desirable student behavior. The observation included a global Likert-style rating of teacher quality for the observer to complete at the end of each observation. It was hypothesized that effective teachers utilize teaching and feedback primarily, and heavily utilize small group instruction. Ineffective teachers hypothetically spend more of their time managing behavior and the environment. Early pilots of the observation have held promise, with adequate sensitivity to teacher behaviors.

The TOT aligns with past research on teacher attributes that correlate with student achievement. For example, Brophy and Good (1986) conducted a qualitative review of over 30 studies investigating teacher-level variables that were observed to improve student level achievement from kindergarten through high school. The authors term these "process-product" variables. Because the research summarized specifically tested for

teacher behaviors that contribute to student success, many of the studies utilized some form of direct observation. Brophy and Good (1986) selected studies that used the teacher as the unit of analysis, averaging student achievement to form a single datum point per class. This typically resulted in a low n for any individual study. Across studies, there were inconsistencies regarding when to question students for comprehension, what types of questions to use, how much control the teacher should have in the classroom, use of reinforcement, and how to proceed through the curriculum. However, a consistent finding across studies was that amount of raw instructional time students were exposed to led to higher achievement.

Students who spent more time in instruction and less time waiting or doing non-academic activities had higher scores on standardized testing. Additionally, students who received brief, prompt, content-specific feedback also tended to have higher achievement. While less consistent across studies, the authors also note that the most successful teachers spent time teaching classroom rules in the beginning of the year as opposed to taking away from instructional time throughout the year to punitively correct students. They summarize this as a "business-like" approach to teaching. Based on this research, the TOT should be valid for quantifying effective teaching by specifically capturing time spent in instruction, time spent giving feedback and the behavioral correction and student wait-time Brophy and Good (1986) refer to.

In the current study, the TOT was used to test the hypothesis that investment in classroom organization early in the year would lead to maximized instructional time later in the year. Observers were in the back of the room during observation and observed behavior of the teacher, not any individual student. The paper based observation was

complemented by an mp3 audio recording that alerted the observer when to observe and record. The audio would provide an auditory cue every 15 seconds with “observation” followed by the number of the interval. Observations occurred three times per classroom per data gathering time period, for a total of 90 minutes of observation across 360 intervals. All observers used this audio recording. In 27.59% of the teacher sample, only two observations could be completed due to limits on available time. A total of 162 observations were completed.

Inter-observer agreement (IOA) was calculated for 19.25% of all observations. IOA was completed by either the author and a graduate student trained in the use of the TOT (training described below) or between two trained graduate students. During IOA sessions, the two observers sat in the back of the room, side by side. An audio splitter was attached to the mp3 player so that there were no time delays. Point-by-point IOA for all observations was 86.87%.

Academic Competence Evaluation Scale

The Academic Competence Evaluation Scales – teacher version - is an 81 item questionnaire covering seven domains of student functioning. These domains load onto two factors: academic enablers and academic skills. It is completed by the general education teacher of the student being evaluated (Diperna et al., 2001; 2005). Academic enablers contain the subscales reading, math, and critical thinking. These subscales reflect teacher perception of a student’s grade level proficiency. Questions are based on individual skills such as spelling and vocabulary. Academic enablers include interpersonal skills, engagement, motivation and study skills. Academic enablers are

skills that precede and are required for students to access and benefit from instruction. Sample questions from these subscales include “participates in class discussion” for engagement or “works effectively in small group activities” for interpersonal skills (DiPerna & Elliott, 2000). The math, reading, motivation to learn, and social skills subscales were used in this study to form a comprehensive picture of student achievement while reducing survey length for teachers.

The ACES uses a five point Likert-style response option. For Academic Skills, a one indicates a skill is “Far Below” age-based norms, while a five would indicate a skill is “Far Above” age-based norms. For Academic Enablers, a one indicates a behavior “Never” occurs, while a five indicates a behavior “Almost Always” occurs (DiPerna & Elliott, 2000). For all questions, a “Not Observed” (marked as “N/O”) can be checked that indicates that the skill cannot be accurately quantified. For missing data, the ACES manual indicated that if two or fewer questions are unrated, they should be assigned the mean value of the scale (a three) and the domain should be scored. The ACES was nationally normed on a geographically and economically diverse sample of 1000 children ranging from Kindergarten through 112th grade. The reading subscale has a reliability of α = .88, math α = .98, interpersonal skills α = .97 and motivation α = .97. Test-retest reliability was also robust: reading was equal to r = .95, math equal to r = .93, interpersonal skills equal to r = .81 and motivation r = .84. The reading subscale strongly correlated to measured reading achievement on the Iowa Test of Basic Skills (r = .80) as did the math subscale (r = .86). The interpersonal skills subscale had moderate convergent validity with the Social Skills Rating System (r = .50; Diperna & Elliott, 2000).

Classroom Practice Measure

RC is comprised of a set of teaching skills, each skill used to address different areas of student and classroom functioning. These skills can be relatively autonomous. Past research has found that teachers often select certain practices within the *RC* model and chose not to employ others (Elliott, 1999). Therefore, implementation of *RC* is not a dichotomous variable, but rather falls along a wide spectrum based on the individual choices of the teacher. The Classroom Practice Measure (CPM) has 34 likert style questions with a scale of zero to five and seven open response questions (Rimm-Kaufman & Chiu, 2007). The measure queries teachers about their use of *RC* in the following domains: hand signals, classroom opening exercises, classroom rules and consequences, classroom organization, introduction to materials, student choice, student reflection, assessment and parent communication, time-out, and use of a problem solving meeting (Rimm-Kaufman & Chiu, 2007). Open response questions are scores zero to five. The CPM contains no language indicating it is an assessment of *RC* fidelity. Attached to the CPM in the current study was a list of demographic questions. This included questions such as “amount of years in higher education studying teaching” and “total years employed as a teacher”.

The CPM has excellent reliability and acceptable validity. Reliability was equal to $\alpha = .94$ in a previous study (Rimm-Kaufman & Chiu, 2007). A sample of 68 *RC* teachers who filled out the CPM showed moderate correlation with two trained observers who went into their classroom ($r = .70$), indicating good concurrent validity (Rimm-Kaufman & Chiu, 2007). Discriminate validity was established by comparing the CPM between

teachers who had *RC* training and those that did not. Summed scores showed a significant difference between the two groups (t = .486, p = .000; Rimm-Kaufman & Chiu, 2007).

The CPM was administered in the fall. However, concerns over the psychometric properties of the CPM between fall and spring data collection led to a reconsideration of its use. Specifically, the CPM did not correlate with any dimensions of the TOT. Correlations ranged from r = .024 to r = .122. This was in contrast to another finding from fall data that exposure to RC professional development did correlate strongly with dimensions of the TOT for the fall. Level of training in *RC*, a question on the demographic survey, was numerically coded, with a zero being no training and *RC* level I and level II training being a six. Ordinal correlations between level of training and dimensions of the TOT were strong: teaching time equal to r_s = -.319, feedback equal to r_s = .448, behavior time equal to r_s = .004, and environment equal to r_s = .415. This pattern of behavior fit with our hypothesis that *RC* teachers would spend more time in environmental management and less time teaching. These conflicting findings raised concerns regarding the CPM's validity in the current study. In addition, it was noted that the wording of the response scale used by the CPM may have threatened validity by introducing response bias.

Due to these concerns, the CPM's Likert-style scale was converted to a three point scale with possible answers being, "No, this is not my present in my class", "Yes, this is present in my class, but not in the way described by Responsive Classroom" and "Yes, this is present in class as defined by the developers of Responsive Classroom" (scored zero, one, and two respectively). To increase teacher participation, open-ended questions were removed. The revised scale used in the current study invalidates the

established psychometric properties of the CPM as discussed in previous literature (e.g., Rimm-Kaufman & Chiu, 2007). Reliability was calculated for the version of the CPM used in the current study, which was equal to $\alpha = .95$. This value indicates excellent reliability. Unfortunately, this statistic is based on an extremely small sample size ($n = 24$), far below what would be expected for a confident estimate of test reliability. The revised CPM used in this study can be viewed in Appendix B.

Timeframe for Surveying

Permission to observe and survey teachers was secured from school principals and district administrators in the spring of 2009. This included every principal in the Western Massachusetts district and deputy superintendent, and the principal of each of the other schools. In the fall, teachers were recruited through e-mail request and shortly thereafter by mailings hand-delivered to school mailboxes. If teachers agreed to participate, they wrote their math and reading block schedules on the letter, which was picked up later by the author. Teachers were given the time of the observation via e-mail at least 24 hours in advance of an observation occurring. Informed consent was secured from all teachers who participated.

Teachers were observed for three 30-minute blocks throughout the first six weeks of each site's academic year. Each set of observations were divided amongst at least two days (typically one hour of observation one day followed by 30 minutes of observation at a later date), although in many cases there was only one observation each day. Shortly thereafter, the CPM and ACES were mailed to teachers or handed to them directly on the last day of observation. Observations and surveying were then repeated in the final six

weeks of the school year in the same fashion. Over the course of the year, three teachers dropped out of the study: one due to a maternity leave, one never completed fall surveys, and one who switched schools within the district between data collection phases.

Observations were completed primarily by the author and a fellow graduate student. The graduate student was in her third year of graduate study, had assisted in developing the TOT, and was very familiar with the research design. A minority of observations were completed by other graduate students from the training program of the author, primarily in their first or second year of course work. Participating graduate students completed a three-hour training on use of the TOT early in the fall and were compensated for travel to schools. The training was not completed until 90% agreement amongst trainees was reached when observing a sample video.

A random number generator was used to select target students for the ACES. Teachers then selected the students when they were organized alphabetically by last name. Teachers proceeded on their own time to secure informed consent from parents of these students using a consent form supplied by the author. If teachers could not get informed consent for a particular randomly selected student, teachers were instructed to select a replacement student of similar behavior topography and academic achievement. Three teachers secured informed consent for only four students. Between the fall and the spring, four students dropped out of the study. The primary reason for this was families moving out of district.

CHAPTER III

RESULTS

The purpose of this study was to see whether observable relationships existed between a teacher's fidelity to *RC*, use of teaching time in the beginning and end of the year, and behavioral and academic student outcomes. To answer these questions, data were split into two levels: student level data (level I) and teacher level data (level II). To understand how teachers chose to use their instructional time, how this relates to the use of *RC*, and whether *RC* modulated the changes in instructional time over the year, correlational analysis were conducted within level I. This investigation was structured as two separate examinations: (a) a correlational analysis of the relationships across dimensions of the TOT and (b) positioning the CPM, TOT, and an interaction between the CPM and pretest TOT data as independent variables (IV), predicting spring level TOT scores.

To understand how instructional practices and fidelity to *RC* affect student outcomes, a series of hierarchical linear models (HLM), also known as multilevel models, were constructed that addressed the predictive power of level II data (TOT and CPM) in explaining level I data (ACES), while controlling for the shared variance caused by groups of students being nested under individual teachers (i.e., within existing classrooms). The rationale for this analysis was that teachers exert a certain amount of common variance on students in their classroom, separate from the students' unique learning characteristics. The purpose of HLM is to quantify and account for this shared variance amongst students and control for it in a regression-based model. This is done by

calculating fixed effect coefficients at the contextual levels (although this can be calculated as random effects as well), that are used to form random-effect intercepts and slopes for the within-group student level data (Raudenbush, Byrk, Cheong, & Congdon, 2004). In the first of these multilevel models discussed presently, CPM scores and difference scores from the TOT dimensions (spring – fall) were used to predict slopes of student growth on measured ACES outcomes. Second and third multilevel analyses focused on how the fall and spring level scores on the TOT independently predicted student achievement. This was done to investigate whether the TOT was sensitive to changes in teaching behavior that relate to varying student outcomes. For each dependent variable (DV), data were cleaned by reviewing the frequency distribution of individual questions from each survey. For all data analysis, missing data were treated with pairwise exclusion. Imputation could not be used because the data were either non-continuous or missing data were not random. This is detailed further as specific variables are discussed.

Teacher Level Results

Descriptive data from the TOT are presented visually in Figure 1. Pre-test means, post-test means, and significance testing for changes over time are presented in Table 2. For all statistical analysis involving TOT data, small- and whole group instruction were combined to form one more broadly defined instruction variable. This was because there was a relative infrequency of small group instruction in the present sample. For the current analysis and all subsequent statistical modeling, the critical p-value was set to .05. Table 2 suggests that feedback significantly increased over the course of the year. Concurrently, time spent managing the environment decreased, the difference

approaching significance. Ignoring student level data, the changes in teaching behavior were modeled with CPM scores as moderating the relationship between fall and spring TOT scores. This was done in block fashion for each dimension of the TOT, with the first block including CPM scores and fall TOT data individually predicting spring scores for each respective spring TOT dimension. The second block included an interaction of CPM scores and fall TOT data.

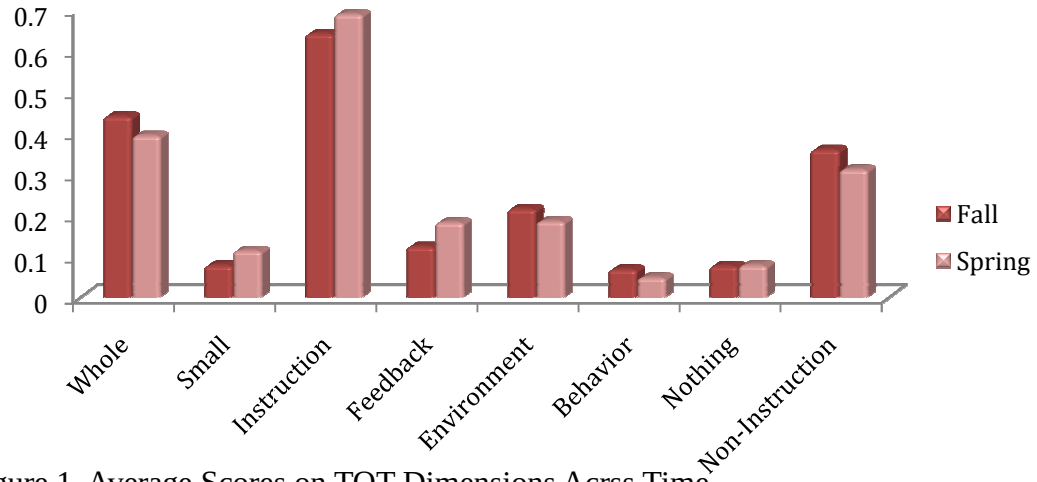

Figure 1. Average Scores on TOT Dimensions Acrss Time

To validate interactions, the model with the interaction should explain more variance than the model with the same predictors included as only individual predictors (Cohen, Cohen, West, & Aiken, 2003). Assumptions of homoscasdiscity were met for feedback and environment, however there was slight heteroscasdiscity for behavior and instruction. Using Q-Q plots, normality of the dependent variables was verified for instructional time, feedback, and environmental management. Time spent delivering

behavior did display a positively skewed distribution, most likely due to its relative infrequency and difficulty in measuring.

Table 2

Average Scores of TOT Dimensions Across Time

	Pretest *M (SD)*	Post-Test *M (SD)*	t-value
Instruction	77.51 (14.00)	83.43 (13.62)	1.50
Feedback	13.36 (13.02)	22.66 (15.63)	2.43*
Environment	26.82 (13.02)	21.85 (8.72)	- 1.72
Behavior	7.19 (5.77)	5.76 (6.40)	- 1.11

n = 23
* significant at the .05 level

CPM scores had a mean of 53.43 (*SD* = 13.03). CPM scores showed evidence of moderate negative skew, indicating that more teachers reported implementing *RC* with high fidelity than with moderate or low fidelity (see Figure 2).

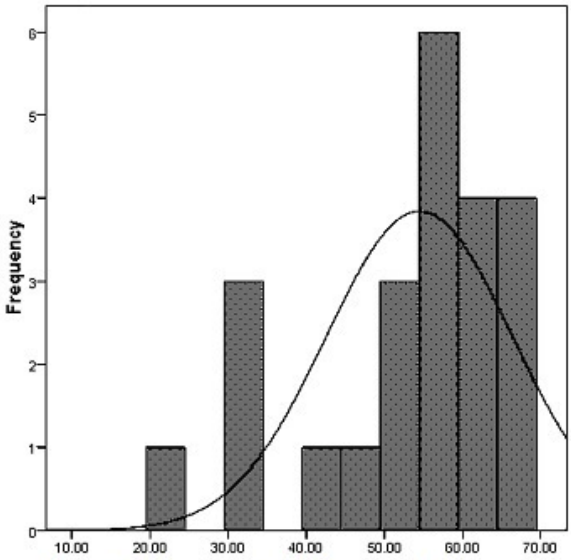

Figure 2. Histogram of CPM scores with estimated curve overlay.

The result of the regression looking at changes in teaching behavior can be seen in Table 3. No linear combination of variables showed to be significant in predicting teaching behavior. The analysis may be underpowered, and behavioral trends may not be linear.

To answer the question of how teachers change their behavior in the beginning and end of the year, a correlational analysis was conducted across dimensions of the TOT in the fall and spring. Results of this analysis are presented in Table 4. Ignoring the phase of data collection, there was a consistent negative correlation between the amount of instructional time delivered and time spent managing the environment and correcting

Table 3
Change in R^2 for Teacher-Level Behavioral Predictors

	Model I R^2	Model II ΔR^2
Instruction	.001	.053
Feedback	.009	.002
Environment	.033	.029
Behavior	.17	.054

* $p < .05$

behavior. There was also a negative relationship between time spent managing the environment and behavioral correction. Looking at temporal variability, there was a negative relationship between time spent correcting behavior in the fall and spring levels of time spent giving feedback to students (r = -.395, p = .028). Fall behavioral correction also predicted spring behavioral correction (r = .463, p = .011). Time devoted to

instruction in the fall predicted spring levels of behavioral correction (r = -.360, p = .042); the more instructional time delivered in the fall, the less time was spent correcting behavior in the spring.

Student Level Results

Descriptive data for the ACES is presented in Table 5. Overall 0.72% of ACES data were missing. This was primarily due to teachers indicating that certain questions were not applicable given the student's current academic level. This made missing data non-random, prohibiting the use of multiple imputation in completing the dataset.

Table 4
Correlation Matrix of TOT Dimensions

	1	2	3	4	5	6	7	8
1. PreIns	-							
2. PreFeed	-.012	-						
3. PreEnv	-.436*	.283	-					
4. PreBeh	-.482**	-.084	-.357*	-				
5. PostIns	.024	.109	.148	-.23	-			
6.PostFeed	.482**	.152	-.255	-.395*	.220			
7. PostEnv	-.116	.063	.193	-.077	-.766**	-.088	-	
8. PostBeh	-.360*	-.149	-.266	.463*	-.416*	-.346	.070	-

* significant at the .05 level (2-tailed)
** significant at the .01 level (2-tailed)

Important to note here is that response options ranged from "Far Below" to "Far Above" for Academic Skills. As such, a student who was rated a "3", or "Grade Level", in the fall and spring would have had a flat slope of improvement over time, yet would have still progressed during the course of the year by staying at grade level. From this

table it can be seen that on average, students showed positive growth in each category across all testing sites, with the exception of interpersonal skills. Site II and site IV showed slight growth on interpersonal skills, while site I and site III showed a slight decrease. Since each site varied widely in regards to size, it would not be statistically sound to compare them using significance testing. However differences across time, collapsed over schools, demonstrated that reading scores ($t = 3.50$, $p \leq .001$), math scores ($t = 11.71$, $p \leq .00$), and motivation scores ($t = 3.36$, $p \leq .001$) significantly increased across the school year. Interpersonal skills decreased slightly over time ($t = 0.09$, $p = .47$).

To address the question of the association of *RC* and teaching behavior on student outcomes, a multilevel regression was done using difference scores (spring – fall) of the TOT dimensions and CPM scores as the predictors of ACES difference scores. Multilevel modeling was appropriate over linear regression as the interclass correlation coefficient (ICC) was considered large (Raudenbush et al., 2004), ranging from $r = .33$ for the motivation subscale to $r = .51$ for reading. HLM 6.08 (Scientific Software International, 2009) was used to calculate results for all models. It was hypothesized that teachers with the largest difference scores of environment (decreases over time) and instruction (increases over time) conform to the "first six weeks" hypothesis and will have the largest degree of positive student change.

P-P plots of the predictors demonstrated that the assumption of normality of the residuals was fulfilled. Q-Q plots showed the same was true of the DV's . Multicolinearity was likely a problem, as an examination of the correlation matrix of the predictors showed significant correlations amongst the dimensions of the TOT. This was

Table 5

Comparison of ACES Scores Across Sample Locations

	Pretest *M (SD)*				Post-Test *M (SD)*			
	Site I	Site II	Site III	Site IV	Site I	Site II	Site III	Site IV
1.	31.25 (14.50)	32.77 (8.64)	34.09 (12.00)	35.97 (11.39)	32.50 (12.92)	35.00 (9.76)	40.27 (8.32)	36.23 (9.95)
2.	24.88 (7.61)	20.82 (4.04)	23.09 (3.65)	22.91 (5.52)	27.63 (9.62)	25.27 (6.39)	29.68 (6.06)	27.11 (7.42)
3.	41.25 (6.69)	41.44 (8.24)	43.27 (5.78)	41.43 (7.09)	40.75 (7.23)	42.47 (7.49)	42.72 (6.01)	41.86 (7.26)
4.	38.00 (13.16)	37.28 (9.69)	42.13 (6.77)	39.95 (9.81)	42.13 (12.11)	38.97 (10.49)	43.79 (7.65)	42.47 (9.56)

1 = Reading, 2 = Math, 3 = Social Skills, 4 = Motivation

not surprising considering the TOT has a fixed amount of intervals; as one variable, such as instruction, increases, another variable, such as environmental management, decreases. For the current analysis, this would not affect the overall proportion of variance explained in the model. However it may create spurious results for the individual coefficients. To reduce multicolinearity, the behavior variable of the TOT was removed for this model and all subsequent analyses. Behavior was observed to be problematic due to its short frequency that may not have been appropriate for momentary time sampling.

Additionally, it was relatively infrequent in comparison to other variables. All IV's were grand-mean centered to further reduce multicolinearity and aid in interpretation.

Results indicated that there was a significant amount of heteroscasdiscity at level I for reading (χ^2 = 46.19, p = .001), math (χ^2 = 51.09, $p \leq .001$) and interpersonal skills (χ^2 = 34.62, p = .042), but not for motivation (χ^2 = 19.474, p >.500). This indicates that students with higher pre-scores in the fall tended to have more variability in scores in the spring, with the exception of motivation. The typical result of a violation of homoscasdiscity is inflation in the standard error associated with each level II coefficient. Typically, this inflation of error in multilevel modeling is slight (Raudenbush et al., 2004). Nonetheless, for DV's with violated assumptions, parallel models were calculated that allowed the heterogeneity of level I variance to be explained by pretest scores (Raudenbush et al., 2004). This revised model did not result in a significantly better model than the original for reading (χ^2 = 2.931, p = .083), math (χ^2 = 1.81, p = .175), nor interpersonal skills (χ^2 = 1.92, p = .162). As such, the original models were maintained. What this suggests is that a potential unaccounted for relevant contextual variable may have existed that was not measured. Other indices of goodness-of-fit suggested this as well (see Table 6). In the following model, variables were defined as:

Level I (student)

$$Y_{\text{ACES}} = \beta_0 + \beta_1 PREACES_1 + r_0$$

In this equation, Y_i is the difference across time of a given ACES domain, β_0 is the intercept and β_1 equals the corresponding prescore of Y_i.

Level II (teacher)

$$\beta_0 = \gamma_{00} + \gamma_{01}INSTRUCTIONDIFF_{01} + \gamma_{02}FEEDBACKDIFF_{02} +$$

$$\gamma_{03}ENVIRONMENTDIFF_{03} + \gamma_{04}CPMSCORE_{04} + \mu_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11}CPMSCORE_{11} + \mu_1$$

In this equation, γ_1 through γ_3 are difference scores from the domains of the TOT. γ_4 was score on the CPM. CPM scores were included at β_1 because this relationship demonstrated that teachers who endorsed the use of *RC* also had a downwards bias in ratings on the ACES. This is a revised model after fit indices demonstrated the misspecification of a former model. The original model included TOT data for the fall as covariates on the difference scores at β_0 and included those same pre-scores on β_1. The hypothesis behind this model was that by the time student ratings occurred, certain instructional effects might have already occurred (fall data collection for students occurred between the 6th and 8th week of the beginning of the school year).

Negative τ intercept values from the unconditional model to subsequent conditional models for math, social skills and motivation provided strong evidence that misspecification had occurred. Snijders and Bosker (1994) stated that misspecification could possibly be the result of non-significant predictors, missing data, or missing variables, among other potential sources, potentially causing anomalous estimates. In the present model, TOT data from the fall was removed as a covariate, as it failed to adequately predict spring level scores, and these same variables were removed from β_1 as the suspected late-rating effect had failed to occur. The lack of other predictors, such as socio-economic status at the student and school level, may also have contributed to the original misspecification.

One indicator of model fit in multilevel modeling is the reduction of error in predicting the DV from one model to another (Snijders & Bosker, 1999). Four different models for this analysis were created for each ACES outcome variable to model how the progressive inclusion of contextual variables explains level I variance both within groups of students and between teachers. The first was the unconditional ICC, which had no predictors. Model A, the null model, included the pre-score of the appropriate DV on level I. Modeling level I completely before comparing the addition of contextual variables is important as cross-level influences across between and within variance can bias individual estimates (Snijders & Bosker, 1994). Model B added context-level TOT difference scores. Finally, Model C was the full model, explained above.

Progressive reduction of error estimates, labeled R_1^2, is shown in Table 6. The addition of pre-scores to the DV of math did not reduce error as was expected, however the addition of TOT difference scores and then CPM scores reduced error predictions. The full model for math accounted for more variance than the unconditional model (an overall difference in between-groups variance of $\tau = .56$), despite unusual patterns in the sequential modeling. Inclusion of TOT difference scores for interpersonal skills resulted in a negative value, indicating misspecification. Further interpretation of results for interpersonal skills in its current form should be met with caution.

Results from the multilevel model are presented in Table 7. Significant differences in student outcomes were noted between teachers for the level I intercept on all dependent variables. For academic skills, the difference scores hypothesized to lead to higher levels of achievement had the opposite effect from the hypothesis. For reading, a significant positive change in instructional time from fall to spring resulted in

Table 6

Modeled Reduction of the ICC for Difference Scores

	Conditional R_1^2 of model A	Conditional R_1^2 of model B	Conditional R_1^2 of model C
Reading	.35	.12	.09
Math	.00	.03	.04
Interpersonal	.18	-.04	.01
Motivation	.16	.02	.03

significantly less teacher-rated reading achievement (γ = -.028, p = .013). An opposite effect was noted for environmental time. This effect approached significance (γ = -0.20, p = .042), as the family-wise error correction reduced the critical p-value. Teacher-rated reading showed a higher slope of progress for students of teachers who endorsed a high level of *RC* use (γ = .80, p = .016). Referring to differences between groups of pre-reading scores, it was observed that teachers who endorsed a high level of *RC* use rated student reading scores in the fall as lower than teachers who endorsed use of *RC* less (γ = -.02, p = .013).

For math a similar pattern was observed as to reading. Large differences in instructional time (γ = -0.09, p = .010) resulted in lower math slopes and differences in environmental management (γ = -.08, p = .010) resulted in higher student growth. The effect of self-endorsed fidelity to *RC* practice approached significance (γ = .47, p = .052). Like reading, teachers who endorsed a higher level of *RC* fidelity displayed a downward bias in ratings, although the effect was not significant (γ = -.02, p = .060). For social

skills only the intercepts were significant, indicating differences between teachers on ratings of social skills. However, the misspecification and lack of significant predictors indicate that the current model is missing important variables.

Finally, CPM scores were significant in predicting teacher ratings of student motivation (γ = .636, p = .003). Like in previous cases, teachers who endorsed a high use of *RC* also rated their students motivational scores lower in the fall (γ = -0.01, p = .002). Unlike the academic domains, differences in instructional time and environmental management between spring and fall did not result in lower teacher-rated scores. Rather, having differences in environment (γ = .08, p = .176), feedback (γ = .08, p = .033) and instruction (γ = .06, p = .341), resulted in non-significant, positive slopes of growth for students.

To test the sensitivity of the TOT and investigate teaching practices more broadly, similar multilevel models were constructed using the fall and spring TOT data as predictors. This was done to test whether varying levels of teaching time at the beginning and end of the school year predicted student ACES scores, ignoring temporal shifts in teaching time that were considered in the previous analysis. The CPM was kept as a predictor since it often accounted for the downward bias in teacher ratings. A test of assumptions revealed similar concerns to the previous model. Normality of the residuals of both levels was adequate for predictors and the dependent variables. Multicolinearity remained a concern for the same reasons mentioned above.

For fall data, the test of heteroscasdiscity was not significant for the motivation subscale (χ^2 = 14.64, p > .500), but was significant for reading (χ^2 = 45.55, p = .001), math (χ^2 = 51.13, p < .001) and interpersonal skills (χ^2 = 34.56, p = .043). Repeating the

heteroscasdiscity correction discussed previously, models with no assumption of homoscasdiscity were calculated. The resulting model for reading was a significant improvement ($\chi^2 = 49.26$, $p \leq .001$) and was chosen over the model that did not control for heteroscasdiscity. The changes to the model were minor, resulting in modifications no greater than the hundredths decimal point for the coefficients. The alternative model for math was not a significant improvement over the original model ($\chi^2 = 2.66$, $p = .099$) nor for interpersonal skills ($\chi^2 = 2.05$, $p = .148$).

For spring level data, significant levels of heteroscasdiscity were observed for reading ($\chi^2 = 47.17$, $p = .001$), math ($\chi^2 = 51.09$, $p \leq .001$) and interpersonal skills ($\chi^2 = 34.62$, $p = .042$), but not for motivation ($\chi^2 = 14.81$, $p > .500$). The appropriate correction was a significant improvement over the original model for reading ($\chi^2 = 25.87$, $p \leq .001$) and the model was modified appropriately. The corrected model was not a significant improvement over the original model for math ($\chi^2 = 2.05$, $p = .148$) nor interpersonal skills ($\chi^2 = 1.46$, $p = .225$). Like in the model of TOT difference predictors, this indicated there is a possible lurking contextual variable. Level I remained the same as in the previous analysis. Level II was constructed as such:

Level II

$$\beta_0 = \gamma_{00} + \gamma_{01}PREINSTRUCTION_{01} + \gamma_{02}PREFEEDBACK_{02} + \gamma_{03}PREENVIRONMENT_{03} + \gamma_{04}CPM_{04} + \mu_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11}CPM_{11} + \mu_1$$

γ_1 through γ_3 represent pre-score observations from the TOT. The CPM was included as a contextual variable on both betas because the prior analysis demonstrated

Table 7
Level II Results of Multilevel Model – TOT Difference Scores

Level II	Reading			Math			Interpersonal Skills			Motivation		
	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio
						β_0						
γ_0	19.29	5.43	3.55*	7.11	2.90	2.45*	15.43	3.25	4.75*	14.12	3.20	4.41*
γ_1	-0.28	0.10	-2.78*	-0.09	0.03	- 2.93*	0.02	0.03	0.58	0.06	0.06	0.98
γ_2	0.03	0.04	0.75	0.02	0.01	1.56	0.02	0.05	0.38	0.08	0.04	2.31
γ_3	-0.20	0.09	-2.20	-0.08	0.03	- 2.94*	0.03	0.05	0.59	0.09	0.06	1.41
γ_4	0.80	0.30	2.68*	0.48	0.23	2.10	0.40	0.21	1.91	0.64	0.18	3.58*
						β_1						
γ_0	-0.48	0.14	-3.33	-0.09	0.12	0.76	-0.37	0.07	-5.03*	-0.30	0.08	-3.98*
γ_1	-0.02	0.01	-2.53*	-0.02	0.01	1.89	-0.01	0.00	-2.11	-0.01	0.00	-3.18*

Note. The Dunn-Bonferroni correction was applied to statistically significant results with academic enablers and academic skills each treated as a family.
Note. Results are weighted by the total number of individual observations done with each teacher.
*significant at the .05 level.

that it served as a biasing factor in how teachers rated student academic skills and academic enablers in the fall.

Goodness-of-fit indices are presented in table 8. For this model, the baseline model (model A) was compared to the unconditional model of no predictors, exactly as with the previous analysis. When fall scores were added as context level predictors, this was compared directly to model A, labeled "B-fall". Fall predictors were then removed, and spring predictors were added in the same fashion, creating the model "B-spring". Note that the fall model and the spring model were never directly compared, as they represented non-nested models. Results from the goodness-of-fit indices show that error in prediction of ACES scores gradually decreased across all models as level II predictors were added. This suggests that across the DV's, the models were an accurate representation of the variance in the DV. Like in the previous analysis, the TOT predictors best represented teacher-rated reading achievement.

For observations completed in the fall, it was observed that teachers who began the year with a strong emphasis on instructional time rated student reading achievement as higher (γ = .28, p = .027) in the spring. This relationship very closely approached significance. At the same time, teachers who emphasized environmental management in the fall also rated overall student reading growth as higher (γ = .11, p = .090). This relationship approached significance. A similar pattern was observed for teacher-rated math achievement. Instructional time in the fall approached significance as a predictor (γ = .09, p = .060) and time spent investing in the environment was significant (γ = .06, p = .004). While not significant, time spent in feedback in the fall had a negative relationship with teacher rated math scores (γ = -.03, p = .371). No distinct pattern of teaching

Table 8
Modeled Reduction of the ICC for Fall and Spring Scores

	Conditional R_1^2 of model A	Conditional R_1^2 of model B-fall	Conditional R_1^2 of model B-spring
Reading	.35	.24	.20
Math	.00	.10	.08
Interpersonal	.18	.07	.07
Motivation	.16	.10	.12

beneficial in the fall neither for interpersonal skills nor for motivation.

A different pattern of optimal teaching was observed for the spring. For reading, no specific teaching category emerged as significant, however time spent in feedback was the only type of teaching that did not have a negative slope (γ = .13, p = .192). A similar pattern was noted for math, however p-values for all categories were very low. Like in the fall, no distinct pattern of optimal teaching emerged for the academic enablers.

Table 9

Level II Results of Multilevel Model – TOT Fall Scores

Level II	Reading			Math			Interpersonal Skills			Motivation		
	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio
						β_0						
γ_0	149.47	5.76	3.38*	6.68	2.79	2.39	15.48	3.20	4.85*	14.49	3.18	4.55*
γ_1	0.28	0.11	2.43	0.09	0.04	2.02	0.02	0.04	0.53	0.00	0.07	-0.01
γ_2	0.04	0.07	0.61	-0.03	0.04	-0.92	-0.04	0.06	-0.60	-0.08	0.07	-1.15
γ_3	0.11	0.06	1.79	0.06	0.02	3.47*	0.01	0.04	0.15	-0.02	0.04	-0.55
γ_4	0.77	0.30	2.52	0.45	0.22	2.10	0.40	0.21	1.96	0.63	0.18	3.49*
						β_1						
γ_0	-0.48	0.15	-3.26*	-0.07	0.11	-0.62	-0.37	0.07	-5.23*	-0.31	0.07	-4.23*
γ_1	-0.02	0.01	-2.53	-0.02	0.01	-1.93	-0.01	0.00	-2.14	-0.01	0.00	-3.12*

Note. The Dunn-Bonferroni correction was applied to statistically significant results with all academic enablers and academic skills each treated as a family.

Note. Results are weighted by the total number of individual observations done with each teacher.

*significant at the .05 level.

Table 10

Level II Results of Multilevel Model – TOT Spring Scores

Level II	Reading			Math			Interpersonal Skills			Motivation		
	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio	Coefficient	St. error	t-ratio
						β_0						
γ_0	19.84	5.80	3.42*	6.98	3.02	2.31	15.28	3.06	5.00*	14.51	3.04	4.78*
γ_1	-0.20	0.17	-1.14	-0.06	0.07	-0.84	0.07	0.06	1.23	0.07	0.09	0.80
γ_2	0.13	0.10	1.36	0.01	0.06	0.23	0.00	0.06	-0.04	0.07	0.06	1.26
γ_3	-0.04	0.21	-0.17	0.02	0.10	0.16	0.09	0.09	0.98	0.19	0.14	1.33
γ_4	0.76	0.31	2.45*	0.46	0.23	1.98	0.40	0.20	1.97	0.65	0.17	3.77
						β_1						
γ_0	-0.49	0.15	-3.33	-0.09	0.13	-0.72	-0.36	0.07	-5.20*	-0.31	0.07	-4.17*
	-0.02	0.01	-2.41	-0.02	0.01	-1.83	-0.01	0.00	-2.15	-0.01	0.00	-3.35*

Note. The Dunn-Bonferroni correction was applied to statistically significant results with all academic enablers and academic skills each treated as a family.

Note. Results are weighted by the total number of individual observations done with each teacher.

*significant at the .05 level.

CHAPTER IV

DISCUSSION

This study tested whether an independent evaluation of *RC* would align with the results of previous studies. To add to the body of literature on *RC*, direct observation was used to test whether teachers who chose to allocate instructional time in the beginning of the year to teach classroom routines and establish behavioral norms had a greater amount of instructional time in the spring, and a corresponding higher slope of student growth, than teachers who kept teaching practices constant from fall to spring. As direct observation has rarely been used as a measured outcome in program evaluation, this study also investigated how varying teaching practices relate to student growth trends via the introduction of the TOT.

Will the Effects of RC Generalize Across Behavioral Constructs Not Previously Measured?

Correlational analysis showed that there were no observable behavioral differences between teachers that reported using *RC* with high fidelity and those that did not. One would expect that use of RC practice would positively correlate with time devoted to environmental management and behavioral correction in the fall and time in instruction and feedback in the spring. This was not the case, with no discernable pattern to observed *RC* teaching practices. It was noted anecdotally in observations that some teachers highly trained in *RC* did spend a large amount of time introducing classroom materials, otherwise known as “guided discovery” (NEFC, 2003). For example, one

teacher who was observed used yellow caution tape to control students handling of classroom materials before they had been introduced in regards to their function and proper use. Another teacher used 20 minutes of a math block to have students brainstorm ways to use certain tools in a math toolkit, then debriefed the class as to their function.

Despite these observations, the TOT was not sensitive to the relative low rate of frequency of these types of behaviors. While it is logical that teachers should be proactive in teaching students the proper use of materials, it appeared *RC* teachers in this sample tended to do this selectively. It is possible that teachers changed their normal course of instruction due to the presence of observers in the room. Furthermore, it could be that teachers were not following the "first six weeks hypothesis" literally. *RC* teachers in normal practice may take only a week or two weeks to do this, which the TOT would largely have missed since observations extended for six weeks. This conclusion was based on relationships within level I data. However, when relationships were investigated across levels, seemingly paradoxical results were observed.

In contrast to level I results, the interactional relationships between teacher-level and student-level data demonstrated that fidelity to *RC* had a significant relationship with select teacher-rated academic achievement. In other words, *RC* training did not explain a significant proportion of variability in teaching behavior, however did explain significant variability in student achievement. Furthermore, certain teaching behavior had a significant relationship with student achievement, or approached significance, in hypothesized directions. A likely reason for this apparent discrepancy is the difference in statistical power between level I (n = 24) and level II (n = 178). As an example, a post-hoc power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2009) revealed

that, given the observed effect between fall levels of environmental management and spring levels of feedback of $r^2 = .065$, power would not reach .70 without the addition of 81 data points. Furthermore, error in the TOT would be compounded when relationships were examined across TOT categories, as opposed to being correlated to a well-validated empirical measure such as the ACES, which may have less error in estimating the true level of behavior.

In this case, review of the TOT could potentially ameliorate this discrepancy. This may include further refinement of operational behavioral definitions based on feedback from this study to increase content validity (Hintze, 2005). Discussed in more detail in the limitations section, the use of momentary time-sampling alone may have been inappropriate for certain categories of behavior. Finally, the effect of situational specificity is likely significant (Kazdin, 1979; Merrell, 2008). Situational specificity is defined as the interaction between the likelihood of a behavior occurring and the behavioral context - the environment may modulate both the temporal frequency and expression of a behavior that may not necessarily generalize to other environments. This creates a confound in determining stability of behavior trends over time and the quantification of within-person reliability, as some of the variance in observed behavior is due to contextual effects (Hintze, 2005). Indeed, teachers move through several distinct contexts during the typical school day, such as math instruction or literacy instruction. The situational demands of these contexts may interact with intra-individual behavioral tendencies, otherwise known as classroom management skills in this case, to modulate the frequency of observed teaching behaviors.

Another plausible explanation is that *RC* as practiced in a naturalistic setting does not result in more instructional time, which results of this study support. The significant relationship between *RC* and academic achievement could be through a more effective use of instructional time, such as use of more effective, direct, teacher language in the classroom during instructional time. While this conflicts with the "first six weeks" theory, it does lend support to *RC* through other aspects of the *RC* theoretical model.

Will Results of an Effectiveness Study Converge With Previous Findings on RC?

Although there were no observable differences in the way *RC* teachers managed their classrooms, there were compelling results at the student level. Students of teachers who endorsed the use of *RC* had a greater slope of progress in teacher ratings of reading and motivation, with the same effect approaching significance for math. Interestingly, this included all student dependent variables except interpersonal skills. In one sense this confirmed past research, such as Rimm-Kaufman et al. (2007), who found that use of *RC* increased math achievement. It also added to the current body of literature by demonstrating a strong relationship between slopes of teacher-rated reading growth, student motivation, and the use of *RC* practices. Elliott (1999) also found significant results for reading achievement. However, this was in contrast to past research, such as Rimm-Kaufman & Chiu (2007), that found non-significant findings for the use of *RC* to improve reading scores. This particular study used simulated grades as a predictor of reading achievement. The discrepancy in results suggests that simulated grades are not a reliable way to quantify reading achievement.

This study's results also conflicted with past findings. Elliott (1999) found in a quasi-experimental between-groups analysis that *RC* did improve teacher-rated social skills of students. This was not the case for this study. It was difficult to isolate a cause of this discrepancy, particularly considering the largest differences were found through teacher report in Elliott (1999); the same method this study employed. The survey instruments used in Elliott (1999), the SSRS, and the ACES in the current study, are only moderately correlated. Furthermore, Elliott's (1999) research was an efficacy study; it was sponsored by the NEFC. It is likely that fidelity to *RC* practice was higher when the developers of *RC* supervised components of the study. Another reason for this apparent discrepancy may be that in certain situations, *RC* is underpowered as an appropriate universal-level prevention program for student social skills.

The discrepancy between the non-significant findings of the observational data and the significant results of the multilevel modeling begs the question as to what are the active ingredients operating in *RC* beyond shifts in teaching time. The authors chose not to look at subscales within the CPM; reliability and validity of individual subscales was unknown. Aside from issues with statistical power, it may also be that *RC* effectiveness is due to one of the many other *RC* tenants that could not be quantified on the TOT. For example, a recent focus of the NEFC has been on the use of teacher language (Denton, 2007). This body of literature stresses the use of "reinforcing, reminding, and redirecting", which borrows from more traditional behavioral theory that has a substantial evidence-base. While this type of language theoretically may allow more instructional time by increasing classroom control, this effect may be too subtle to observe with the measurement instruments used in this study.

RC also emphasizes constant parent communication (NEFC, 2003). This link between the home and school, unobservable in the classroom, may be a significant influence on the current results. Interestingly, there is a paucity of contemporary research that specifically examines how parent participation in classroom activities relate to student academic outcomes. A descriptive analysis of parent involvement in schools conducted by Zill and Nord (1994) reported that the populace of parents who both attended a general school meeting, attended at least one school event, and volunteered for at least one function each school year had less than half the percentage of children in the bottom half of the class academically than parents who reported doing none of these things (26% and 56% respectively).

Fanutzzo, McWayne and Perry (2004) established relationships between parent involvement and student outcomes in a sample of 144 students enrolled in pre-school Head Start. Family Involvement was assessed by surveying parents using the *Family Involvement Questionnaire* (Fantuzzo, Tighe, & Childs, 2000), which is comprised of three major factors: school-based involvement, home-based involvement, and conferencing. Student skills were assessed by surveying teachers using the *Preschool Learning Behavior Scale* (McDermott, Green, Francis, & Scott, 1996) and the *Conners Teacher Rating Scale-28* (Conners, 1990). Student achievement was measured with the *Peabody Picture Vocabulary Test* (3rd ed.; Dunn & Dunn, 1997). This study found significant positive relationships between the school-based parent involvement construct and receptive vocabulary skills (r = .32) and academic skills (r = .23 to r = .25). A significant negative correlation was found between school-based parent involvement and teacher reported inattention (r = -.20) and conduct problems (r = -.29). Even stronger

findings were reported for home-based family involvement. Students within an *RC* classroom may be reading more at home or completing more homework assignments, in addition to allowing more functional communication pathways between the teacher and parents. This might allow more generalizable behavior modification from the classroom to the home.

Finally, it may be that certain elements of *RC* negate each other in regards to observed instructional time. For example, while *RC* encourages that teachers use teaching time to reinforce the social curriculum in the beginning of the year, it also stresses the maintenance of a highly organized, efficient classroom (NEFC, 2003). The former would reduce instructional time; the latter however might increase it.

How Does Initial Investment in Classroom Organization Play Out as a Cost-Benefit Analysis Across the School Year?

The data suggested that "the first six weeks" theory is not the optimal way to manage a classroom. On the contrary, having large discrepancies in instructional time significantly reduced teacher rated reading and math growth. Furthermore in the analysis of teaching behavior only, there was a significant negative correlation between instruction time in the fall and spring behavioral corrections. In other words, teachers who had low instructional time in the fall had a more difficult time controlling their classroom in the spring. This suggested that strong teachers immediately put emphasis on instructional time. Having students engaged during instructional time reduced the potential for behavior problems over the course of the year. This was further supported

by the observation that teachers who spent a significant amount of time correcting student behavior in the fall continued to do so in spring.

This observation may also be influenced by the pattern of growth for student motivation. While not significant, there was a positive relationship between feedback time in the spring and teacher ratings of student motivation. Teachers who start with a strong emphasis on instruction, moving into feedback in the spring, may have students who feel more accomplished than students with teachers who spent a majority of time working only on classroom environment. This self-efficacy may increase student motivation, which could reduce behavioral problems. This hypothesis is supported by extant research. For example, Chen (2003) conducted a path analysis that measured the relationship between math achievement, student self-efficacy, student's self-evaluation, and self-judgments regarding effort put into academics. The sample included 107 seventh grade general education students. Math was measured with the ITBS and self-efficacy was measured with a math-specific self-efficacy measure designed by the authors. It aligned with the ITBS and asked questions regarding how confident students were they could answer certain questions on a Likert scale. A correlation of $r = .50$ ($\beta = .50$, $p < .05$) was found between self-reported self-efficacy and concurrently measured ITBS math scores.

The "first six weeks" hypothesis did show partial support. The multilevel model of the difference scores showed that teachers with large differences in environmental management did rate student achievement higher in math (significant) and reading (non-significant). In summary, teachers who put a strong emphasis on environmental management *and* maintained high levels of instructional time had students with the

steepest slope of growth over the year. Teachers who maintained this high level of environmental management over the course of the year lost the investment's beneficial effect. This may be reflective of an inability to have students self-regulate their own behavior through effective modeling, correction, and reinforcement. This type of classroom management would lend itself well to independent work that was often marked as feedback time on the TOT for this study.

Taken together, the findings suggested that training in RC does improve a teacher's ability to motivate their students and improves teacher perceptions of reading and possibly math achievement. *RC* practices, such as maintenance of an orderly classroom, goal setting, positive, specific teacher language, and parent involvement may combine to result in a significant positive effect for students. However, this study also demonstrated that a heavy emphasis on tasks outside of curricular instruction alone, such as teaching classroom routines, in the beginning of the school year is not best practice if it takes away from instructional time. While morning meetings, heavily emphasized in RC, may be important, practices such as this must not significantly take away from the instructional time students need.

Can Instructional Behavior of the Teacher be Observed and Quantified in a Reliable and Valid Manner?

This study introduced the TOT, a measure designed to observe teaching behavior in the classroom. Throughout this study, the TOT was a robust and reliable measure of teaching behavior. Furthermore, the TOT revealed an optimal pattern of teaching behavior over the course of the year that resembles theory on the gradual release of

responsibility instructional framework (Pearson & Gallagher, 1986). In the gradual release of responsibility framework, teachers first tightly control student learning by modeling lesson goals and strategies for knowledge acquisition (i.e., “I do it”; Fisher & Frey, 2008). Teachers then engage in collaborative instruction with students using various levels of scaffolding (i.e., “we do it”). Pearson and Gallagher (1986), in their literature review of explicit strategies to teach reading comprehension, refer to this as “guided instruction”. In their conceptualization, teaching is gradually scaled down and control of application of recently learned skills is given to the students to facilitate application and automaticity. This is followed by higher order comprehension and synthesis when students use their knowledge to make connections and inferences from the lesson in peer-to-peer collaborative work or independent work (i.e., “you do it”; Fisher & Frey, 2008).

The gradual release of responsibility follows a repetitive pattern from unit to unit of the curriculum. However, Fisher and Frey (2008) stated that as the academic year progresses and students build independent work skills and grade level comprehension of material, collaborative or independent work would become more frequent in general. It is at this time that teacher feedback would also become more frequent as teachers shift from whole-class instruction to providing tailored feedback to individual students and small groups of students. In other words, students learn both academic content and learning management skills at the beginning of the school year. As the year progresses, students can apply these self-organizational skills to manage their own learning given a clear objective, allowing for different types of instruction (Fisher & Frey, 2008). For example, fall instructional time approached significance as a predictor of teacher-rated reading

achievement. As instructional time went up in the fall, so did reading scores. However in the spring, instructional time had the opposite effect; the more instructional being given, the lower the rating of reading and math scores. Time spent giving students individualized feedback had taken its place as the beneficial variable, although this relationship was not significant.

This aligns with Bloom's seminal theoretical work on different types of student learning that results in higher levels of mastery learning (Bloom, 1965). A student whose learning is carefully controlled through teacher-guided instruction may develop a level of knowledge equivalent to Bloom's level of "comprehension", which allows one to state a fact or rule in one's own words. However, a student who can perform independent work with teacher feedback may reach a higher level of comprehension, such as Bloom's highest level of learning, "synthesis", which requires the integration of multiple elements to create new meaning. A plausible theory, requiring more evidence, is that teachers who devote time to both instruction and environmental management in the fall not only maximize instruction, but open the gates for higher levels of learning later in the year.

RC does not specifically endorse the use of gradual release of responsibility regarding instruction. However, *RC* does promote this specific framework for the development of appropriate social behaviors and introduction of class materials (Crowe, 2009). It is possible that teachers who learn about the gradual release of responsibility through development of classroom management strategies and the build-up of positive student behaviors through *RC* professional development also learn to generalize the theory to academic instruction. This general release of behavioral skills to students may result in strong self-regulatory skills for students, making class time more efficient with

less time spent managing instructional transitions. If such is the case, the *RC* teacher may insert the social curriculum into the classroom throughout the year using the residual time saved through this release of responsibility. In other words, the *RC* teacher may be able to address problems across both academic and behavioral dimensions with no net loss to instruction.

It is important to note that giving feedback does not necessarily indicate that the release of responsibility has been adequately executed. Teachers with poor instructional strategies may skip directly from introducing a lesson to having students work independently on connected material. Because no scaffolds were provided, students may not achieve an optimal level of comprehension. The TOT is limited in this sense because it only captures the immediate frequency of feedback provided to students. It does not provide conditional frequencies based on the prior release of responsibility. An improved teaching quality scale is one way to resolve this issue, as well as increased observations for each teacher.

The TOT also showed discriminant validity by predicting achievement only in areas one would hypothesize instructional time to have an effect. TOT dimension scores only had significant relationships with academic skills, not academic enablers. This was true for TOT results from both the fall and spring. The exception to this general pattern was a non-significant positive slope of growth in motivation for teachers with large differences in feedback scores from fall to spring. This once again may reflect a release of responsibility approach to teaching. Students may have stronger academic self-efficacy when they can understand and synthesize information independently.

Taken together, results of this study suggest that *RC* is effective in increasing teacher perceptions of student reading achievement and student motivation, with a positive effect for mathematics approaching significance. In contrast to past research, *RC* had no significant effect on the development of social skills. The "first six weeks" hypothesis was only partially supported. Teachers had to both maintain high levels of environmental management *and* instructional time in the beginning of the year for the behavioral investment to be effective. As "the first six weeks" hypothesis would predict, teachers who could not release responsibility to students over time had lower student achievement than teachers that could transition out of environmental management by the conclusion of the year.

Limitations

The ACES is not a direct measure of student ability in regards to academic achievement. It is a proxy – an opinion – completed by the teacher. Given the logistical constraints of this study, direct assessment of student achievement was unduly prohibitive. The same concern arose for use of the CPM. While the CPM demonstrated validity in this study and previous research, it is a proxy of actual fidelity to RC. It also is important to reiterate that due to modifications of the CPM response scale in the current study, previously established psychometrics of the CPM may be invalid.

As with any study based on correlational analysis, there is never certainty regarding causation. While the relationship between *RC* and student outcome variables were strong, it was possible that a lurking variable created the illusion of a direct relationship. For example, *RC* teachers may rate student ability in the spring higher than

it actually is. This may be due to dissonance between belief about the effectiveness of *RC* and actual student performance, resulting in beliefs that the *RC* method must help students improve. Participant bias in the study is also a potential threat to validity. Although never explicitly stated by the researchers, the rescaling of the CPM alerted teachers that the current study was focused on *RC*. They may have been enthusiastic to show the program they had invested so much time in was effective, inflating spring level scores.

Goodness-of-fit indices suggested that important contextually based variables were missing from the analysis. Given past research that has used multilevel modeling on a school based-population (e.g., Roberts, Mohammed, & Vaughn, 2010), it is very likely that at least one of these variables is a measure of economic status, both at the school level and student level. This study also looked at classrooms that were average in many ways. For example, class size was average, ranging from 8 to 22 students, and most teachers had masters-level training. Classroom characteristics nearer the extremes, such as having a very large class size, may interact with target IV's to produce different patterns of student achievement.

While this study added to the generalizability and effectiveness of *RC*, there were methodological concerns that should be considered. The sample size of teachers was low. Results drawn from only teacher level data are potentially underpowered, increasing the chance of type II error. Previously mentioned, CPM scores were negatively skewed. As the primary measure of *RC* fidelity, this skewed distribution may have resulted in a loss of overall validity for the study. There were several potential causes for this skew. One, this study took place in an area close to the location of the NEFC. This influence may

have resulted in a high concentration of teachers exposed to *RC*. Second, construct validity may have been threatened by social desirability bias. A third possible cause was that the CPM does not properly differentiate *RC* from other types of teaching behavior, an issue of discriminant validity, despite previous findings (e.g., Rimm-Kaufman et al., 2007).

Finally, due to logistical limitations (e.g., number of observers), we could not observe teachers for more than a total of one and a half hours each semester. This potentially increased the chance of type II error, since certain behaviors may have been infrequent, requiring a much longer observation period to reliably detect. This may have resulted in a decrease of overall generalizability of scores. Previous research would suggest this magnitude of observational time is less than ideal in reliably quantifying a priori' defined target behaviors of a given subject (Hintze & Matthews, 2004). This concern raises the question as to whether teaching behaviors demonstrate similar stability characteristics to student behavior. In other words, what magnitude of observation time is necessary to conclude behavior has been reliably quantified? As systematic direct observation of teaching behavior increases in frequency, for consultative purposes for example, this question becomes of critical importance.

Along these same lines, the two data gathering periods likely did not fully capture longitudinal patterns of shifting teaching behavior. To better model hypothesis of effective teaching, such as the gradual release of responsibility, more observational periods should be included throughout the academic year. This also would provide the requisite information to rule out other longitudinal growth patterns aside from a linear

trajectory, such as quadratic or cubic trends. If curvilinear trends exist, this would confound the current results.

Directions for Future Research

This study introduced the TOT as an outcome measure that isolates one of the medial cogs between professorial development and student outcomes, changes in instructional time in class. In the current study the TOT predicted student-level academic achievement. However, future research needs to be directed towards understanding the convergent and discriminant validity of the observation tool, and its generalizability across different times of the day, different subjects, and teachers of different age groups. Such analysis is necessary to build a foundation of empirical evidence for the use of TOT as an applied instrument

Current results also suggest the TOT needs to be modified. For example, turning the time spent correcting behavior code into a frequency count so that it can be more accurately quantified. Differentiating between time spent encouraging positive behavior, such as giving praise, and time spent reducing negative behavior, such as correcting student behavior or redirecting students, can forward the reliability and construct validity of the TOT. Refining the measure and better understanding its psychometric properties will advance the TOT as an important consultative tool for psychologists and administrators.

While a brief likert-rating accompanied the TOT to document quality of teaching, this is neither reliable nor valid enough for research or applied use. Quantifying quality of teaching is important, as any of the defined categories of the TOT could be done by a

teacher poorly (although this would likely result in excessive behavioral and environmental management as control of the learning environment is lost), creating a confound based on these scores. Revision of the quality scale may include extending or restructuring its response scale, creating more clear definitions, and increasing training time for observers.

The downward bias in fall level ratings for teachers with higher CPM scores introduced another research question, "do *RC* teachers have a more accurate understanding of their students' academic and social abilities?" Future studies may want to empirically test this hypothesis for validation by correlating CPM scores to both direct and indirect measures of student ability. *RC* teachers may pre-assess students more in the beginning of the year, be more sensitive to student deficits, or simply may be more cynical of student performance in the beginning of the year.

This study tested the hypothesis that "the first six weeks" is an effective strategy to maximize student growth. *RC* is a constellation of teaching practices, only one of which is the "first six weeks." A more comprehensive cost-benefit analysis of this strategy, and the many others, such as morning meeting, logical consequences, or guided discovery, is needed. Future research experimenting with different combinations of strategies may be able to isolate the independent effectiveness of other strategies within the *RC* program. Along these lines, there are likely optimal levels of time spent delivering strategies nested within the *RC* social-curriculum. Time spent in these strategies might have a curvilinear relationship with academic achievement, such that a classroom management strategy or social skills lesson may be beneficial until it significantly takes

away from instructional time (Brophy & Good, 1986). What these optimal times are, and what constitutes significant time for instruction, remains to be discovered.

Finally, results from this study showed that the teaching strategies that may benefit student motivation and student interpersonal skills may not benefit student reading and math achievement. In this study, teaching strategies that resulted in growth for reading and math did not necessarily result in improvements for academic enablers, and vice versa. While previous research has demonstrated that social skills development can result in achievement gains (e.g., Caprara et al., 2000; Malecki & Elliott, 2002; Wentzel, 1993), there is much to learn regarding how certain instructional and curricular choices differentially impact academic and behavioral growth. Direct observation can be used as one outcome measure to further explore this complex system of inter-relationships.

APPENDIX A

TEACHING OBSERVATION TOOL

Teaching Observation Tool

Date of Observation: _____ Time: _____

Grade Level: _____ # of Adults: _____ # of Students: _____

Content of the Lesson: Reading Writing Math Social Studies Science Other: _____

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	31	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	41	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	51	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	61	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	71	72	73	74	75	76	77	78	79	80
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	81	82	83	84	85	86	87	88	89	90
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	91	92	93	94	95	96	97	98	99	100
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	101	102	103	104	105	106	107	108	109	110
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	111	112	113	114	115	116	117	118	119	120
Teaching	T	T	T	T	T	T	T	T	T	T
	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W	S / W
Feedback	F	F	F	F	F	F	F	F	F	F
Environment	E	E	E	E	E	E	E	E	E	E
Behavior	B	B	B	B	B	B	B	B	B	B

	Poor	Weak	Typical	Good	Excellent
Teaching Quality	1	2	3	4	5

Did students Transition? Yes No
If Yes, what did they transition to? _____

_____

APPENDIX B

CLASSROOM PRACTICE MEASURE

Classroom Practice Measure
***Revised by Benjamin Solomon, Spring 2010**[1]

TEACHER NAME:_____

0

Teachers have many responsibilities and tasks to perform in a very short day. We know that not everything can be a top priority in your classroom, but that you must pick and choose where to focus your energy. Every classroom teacher varies in the practices he or she uses. Through this questionnaire, we would like to learn more about the practices that you use. Please remember that there are no right or wrong answers, but that we are using this questionnaire to describe your classroom.

Directions: Please place a check mark in the box that best describes your use of the stated practice. If you don't know how Responsive Classroom describes the use of a practice, but you believe you do it, mark "Yes, this is present in my class, but not in the way described by Responsive Classroom."

	No, this is not present in my class.	Yes, this is present in my class, but not in the way described by Responsive Classroom.	Yes, this is present in my class as defined by the developers of Responsive Classroom.
1. I use hand signals consistently (e.g. raises a hand for the class to "quiet down")			
2. The class sits in a circle during morning meeting.			
3. Morning meeting includes activities to build community and promote academic and social learning.			
4. Students are greeted during, or before, morning meeting.			
5. Classmates usually greet each other during morning meeting.			

[1] Original measure authored by Rimm-Kaufman & Sawyer, 2001, UVA Social Development Laboratory.

	No, this is not present in my class.	Yes, this is present in my class, but not in the way described by Responsive Classroom.	Yes, this is present in my class as defined by the developers of Responsive Classroom.
6. There is a specific time set aside for children to share things, events, and feelings about themselves.			
7. Morning meeting usually include an age-appropriate playful activity with social and academic purpose.			
8. There is an age appropriate prepared message on the chart/blackboard that contains the important news for the day.			
9. I work together with students to form classroom rules and expectations.			
10. When a rule or consequence is introduced, modeling and discussion are used to emphasize the appropriate behavior.			
11. Rules are stated in the positive (e.g. "We will take care of our classroom by cleaning up our materials".			
12. Rules are posted at the students' eye level in a conspicuous place in the room			
13. I remind, reinforce, and redirect children in the practice of the rules when appropriate.			

	No, this is not present in my class.	Yes, this is present in my class, but not in the way described by Responsive Classroom.	Yes, this is present in my class as defined by the developers of Responsive Classroom.
14. Logical consequences are used to hold children accountable for the rules (e.g. If the child makes a mess, he or she must help clean up. If the child hurts someone's feelings, he or she must make amends.).			
15. The classroom is orderly, accessible, and allows for safe and independent work.			
16. The classroom is arranged in a manner to accommodate whole group, small group, and individual work.			
17. There is an established class meeting space.			
18. The majority of bulletin boards display students' work.			
19. New materials are methodically introduced before making them available for student use.			
20. Students demonstrate ideas for use of materials, try them out with supervision, and think through their care.			
21. Students use classroom materials in choice and independent work regularly.			

	No, this is not present in my class.	Yes, this is present in my class, but not in the way described by Responsive Classroom.	Yes, this is present in my class as defined by the developers of Responsive Classroom.
22. Materials are complete and are stored accessibly for students.			
23. Children's development is a major consideration in choosing lessons.			
24. Academic work is a balance of teacher-directed and student-initiated learning. Most days students experience choices in how they do academic work.			
25. Students share their work with other students regularly.			
26. Teachers, parents, and students collaborate to set goals for the students.			
27. Children's progress is documented in a portfolio type assessment throughout the year. 28. Students reflect on their work and participate in self-evaluation in addition to the teacher's assessment.			
29. There is an established process for children who need quiet time, (e.g. time-out)			
30. "Time-out" is used as an opportunity to help students regain control.			

	No, this is not present in my class.	Yes, this is present in my class, but not in the way described by Responsive Classroom.	Yes, this is present in my class as defined by the developers of Responsive Classroom.
31. Structures are in place (and are used) that enable students to participate in solving classroom problems (e.g., problem solving class meetings, conflict resolution strategies, social conferences).			
32. Students know that they can set the agenda for a class meeting and they know how to do this.			
	No, this is not present in my school.	**Yes, this is present in my school.**	[redacted]
33. Teachers know most of the other teachers' names at the school.			
34. There is frequent and comfortable interaction between the professionals (i.e., teachers) and the paraprofessionals (i.e., administrative assistants, lunchroom staff) at the school.			
34. The school emphasizes the importance of teacher-teacher collaboration.			
35. Teachers collaborate with other teachers at the school to solve problems and develop activities.			
36. Positive structures in the school are in place for teachers to help students.			

APPENDIX C

LIST OF ACRONYMS

ACES	Academic Competence Evaluation Scales
ALT	Academic Learning Time
ANCOVA	Analysis of Covariance
APA	American Psychological Association
AYP	Annual Yearly Progress
BTES	Beginning Teacher Evaluation Study
CASEL	Collaborative for Academic, Social, and Emotional Learning
CFI	Comparative Fit Index
CMT	Connecticut Mastery Test
CPM	Classroom Practice Measure
CRT	Context Relevance Theory
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DRP	Degrees of Reading Power Test
DV	Dependent Variable
ESP	Early Screening Project
GPA	Grade Point Average
GST	General Systems theory
HLM	Hierarchical Linear Modeling
ICC	Intraclass Correlation
IOA	Interobserver Agreement
ITBS	Iowa Test of Basic Skills
IV	Independent Variable

NEFC	Northeast Foundation for Children
NNFI	Non-Normed Fit Index
ODR	Office Discipline Report
PBS	Positive Behavioral Support
RC	Responsive Classroom
RTI	Response to Intervention
SD	Standard Deviation
SSBD	Systematic Screening for Behavioral Disorders
SST	Social Skills Training
STBS	Stanford Test of Basic Skills
STRS	Student Teacher Relationship Skills
SSRS	Social Skills Rating System
SWPBS	School-Wide Positive Behavioral Support
TOT	Teaching Observation Tool
WRMT-R	Woodcock Reading Mastery Test, Revised

REFERENCES

Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory.* Edgewood Cliffs: Prentice-Hall.

Bertalanffy, L.V. (1968). *General Systems Theory: Foundations, Development, Applications (Revised Edition).* New York: George Braziller.

Biglan, A., Brennan, P.A., Foster, S.L., & Holder, H.D. (2004). *Helping Adolescents at Risk: Prevention of Multiple Problem Behaviors*. New York, NY: The Guilford Press.

Bloom, B.S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain* (2nd ed.). Boston: Addison Wesley Publishing Company.

Bradley, R., Hallahan, D.P., & Danielson, L. (Eds.) (2002). *Identification of learning disabilities: Research to policy*. New York, NY: Taylor & Francis.

Bray, M.A., Kehle, T.J., & Hintze, J.M. (1998). Profile analysis with the Weschler Scales: Why does it persist*? School Psychology International, 19,* 209-220.

Briesch, A.M., Chafouleas, S.M., & Riley-Tillman, T.C. (2010). Generalizability and Dependability of Behavior Assessment Methods to Estimate Academic Engagement: A Comparison of Systematic Direct Observation and Direct Behavior Rating. *School Psychology Review*, 39(3), 408-412.

Bronfenbrenner, U., & Morris, P.A. (2006). The Bioecological Model of Human Development. In W. Damon, & R. M. Lerner (Vol. eds.). *Handbook of Child Psychology: Vol. 1. Theoretical models of human development* (pp. 993-1028). New York, NY: Wiley and Sons.

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-375). New York, NY: Macmillan.

Caldarella, P., & Merrell, K. (1997). Common dimensions of social skills of children and adolescents: A taxonomy of positive behaviors. *School Psychology Review*, *26*, 264-284.

Caprara, G.V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P.G. (2000). Prosocial Foundations of Children's Academic Achievement. *Psychological Science, 11*, 302-306.

Carr, E.G., Horner, R.H., Turnbull, A.P., Marquis, J.G., McLaughlin, D.M., McAtee, M.L., et al. (1999). *Positive Behavior Support for People With Developmental*

Disabilities: A Research Synthesis. Washington, DC: American Association on Mental Retardation.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 65, 723-733.

CASEL (2009). Safe and Sound: An Educational Leaders Guide to Evidence-Based Social and Emotional Learning (SEL) Programs. Retrieved from the CASEL website: http://www.casel.org/downloads/Safe%20and%20Sound/1A_Safe_&_Sound.pdf.

Chen, P.P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences*, 14, 79-92.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Conduct Problems Prevention Research Group (1999). Initial Impact of the Fast Track Prevention Trial for Conduct Problems: I. The High-Risk Sample. *Journal of Consulting and Clinical Psychology, 67*, 631-646.

Connecticut State Department of Education (2006). *Connecticut Mastery Test* (4th ed.). Author.

Connor, C.M., Morrison, F.J ., & Katch, L.E. (2004). Beyond the Reading Wars: Exploring the Effect of Child-Instruction Interactions on Growth in Early Reading. *Scientific Studies in Reading*, 8(4), 305-336.

Connors, C.K. (1990). *Conners Rating Scales manual*. Toronoto, Canada: Multi-Health Systems.

Crowe, C. (2009). *Solving Thorny Behavior Problems: How Teachers and Students Can Work Together.* Turner Falls, MA: Northeast Foundation for Children.

Denton, P. (2007). The *Power of Our Words: Teacher Language that Helps Children Learn*. Turner Falls, MA: Northeast Foundation for Children, Inc.

DiPerna, J.C., & Elliott, S.N. (2000). *Academic competence evaluation scales*. San Antonio, TX: The Psychological Corporation.

DiPerna, J.C., Volpe, R.J., & Elliott, S.N. (2001). A Model of Academic Enablers and Elementary Reading/Language Arts Achievement. *School Psychology Review, 31*, 298-312.

DiPerna, J.C., Volpe, R.J., & Elliott, S.N. (2005). A model of academic enablers and mathematics achievement in the elementary grades. *Journal of School Psychology, 43,* 379-392.

Dodge, K.A., & Sherrill, M.R. (2007). The Interaction of Nature and Nurture in Antisocial Behavior. In D.J. Flannery, A.T. Vazsonyi, & I.D. Waldman (Vol eds.). *The Cambridge Handbook of Violent Behavior and Aggression* (pp. 215-242). New York, NY: Cambridge University Press.

Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Services.

Elliott, D.S., Hamburg, B.A., & Williams, K.R. (Eds.) (1998). *Violence in American schools: A new perspective*. New York, NY: Cambridge University Press.

Elliott, S.N. (1999). A Multi-Year Evaluation of the Responsive Classroom Approach: Its Effectiveness and Acceptability in Promoting Social and Academic Gain. Springfield, MA: Northeast Foundation for Children & Kensington Avenue Elementary School Staff. Retrieved January 25th, 2008, from http://www. responsive classroom.org/ about /research.Html.

Elliott, S.N., DiPerna, J.C., Mroch, A.A., & Lang, S.C. (2004). Prevalence and Patterns of Academic Enabling Behaviors: An Analysis of Teachers' and Students' Ratings for a National Sample of Students. *School Psychology Review, 33,* 302-309.

Erchul, W.P., & Martens, B.K. (2002). *School Consultation: Conceptual and Empirical Bases of Practice*. New York, NY: Springer Publications.

Fantuzzo, J., McWayne, C., & Perry, M.A. (2004). Multiple Dimensions of Family Involvement and Their Relations to Behavioral and Learning Competencies for Urban, Low-Income Children. *School Psychology Review,* 33(4), 467-480.

Fantuzzo, J., Tighe, E., & Childs, S. (2000). Family Involvement Questionnaire: A multivariate assessment of family participation in early childhood education. *Journal of Educational Psychology*, 92, 367-376.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2009). *G*Power* (Version 3.1.2) (Computer Software).

Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J. et al. (1978). *Teaching behaviors, academic learning time, and student achievement: a final report of Phase III-B of the Beginning Teacher Evaluation Study. Technical Report V-1.* San Francisco, CA: Far West Laboratory for Educational Research and Development.

Fisher, D., & Frey, N. (2008). *Better Learning Through Structured Teaching: A Framework for the Gradual Release of Responsibility*. Alexandria: Association for Supervision and Curriculum Development.

Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., Gottfredson, D., Kellam, S., Moscicki, E.K., Schinke, S., Valentine, J.C., & Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science, 6*, 151 – 175.

Fletcher, J.M., Fuchs, L.S., Lyon, G.R., & Barnes, M.A. (Eds.) (2006). *Learning disabilities: From identification to intervention*. New York, NY: Guilford Press.

Ford, D.H. & Lerner, R.M. (1992). *Developmental Systems theory: An Integrative Approach.* Thousand Oaks, CA: SAGE Publications.

Fuchs, L.S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13(4), 204-219.

Fuchs, D., & Fuchs, L.S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93-99.

Fuchs, D., Mock, D., Morgan, P.L., & Young, C.L. (2003). Responsiveness-to-Intervention: Definitions, Evidence, and Implications for the Learning Disabilities Construct. *Learning Disabilities Research and Practice, 18*, 157-171.

Gibson, V., & Hasbrouck, J. (2007). *Differentiated Instruction.* New York, NY: McGraw-Hill.

Good, R.H., & Kaminski, R.A. (eds.) (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement. Available at http://dibels.uoregon.edu

Gottfredson, G.D., Gottfredson, D.C., Czeh, E.R., Cantor, D., Crosse, S.B., Westat, I. H. (2000). *Summary: National study of Delinquency Prevention in Schools* (Available from Gottfredson Associates, Inc. 3239B Corporate Court, Ellicott City, Maryland 21042).

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

Gresham, F.M., Elliott, S.N., & Evans-Fernandez, S.E. (1993). *Student self-concept scale.* Circle Pines, MN: American Guidance Service.

Gresham, F.M., Sugai, G., & Horner, R.H. (2001). Interpreting Outcomes of Social Skills Training for Students with High-Incidence Disabilities. *Exceptional Children, 67,* 331-341.

Gresham, F.M., & Project REACH (2005). Response to Intervention: An Alternative Means of Identifying Students as Emotionally Disturbed. *Education and Treatment of Children*, *28*, 328-344.

Grossman, H.J. (Ed.) (1983*). Classification in mental retardation*. Washington, DC: American Association on Mental Deficiency.

Harcourt, Brace. (1987). *Standard Achievement Test*. Orlando, FL: Author.

Hintze, J.M. (2005). Psychometrics of Direct Observation. *School Psychology Review*, 34(4), 507-519.

Hintze, J.M., & Matthews, W.J. (2004). The Generalizability of Systematic Direct Observations Across Time ad Setting: A Preliminary Investigation of the Psychometrics of Behavioral Observation. *School Psychology Review*, 33(2), 258-279.

HLM for Windows (Version 6.08) (Computer software). Lincolnwood, IL: Scientific Software International.

Hoagwood, K., & Erwin, H.D. (1997). Effectiveness of school-based mental health services for children: A 10-year research review. *Journal of Child and Family Studies*, *6*, 435-452.

Hoover, H., Hieronymus, A., Frisbie, D., & Dunbar, S. (1993). *Iowa Test of Basic Skills.* Chicago, OH: Riverside.

Institute of Educational Science (2009). *What Works Clearinghouse.* Retrieved from http://ies.ed.gov/ncee/wwc.

Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., Youngstrom, E. (2001). Emotion Knowledge As a Predictor of Social Behavior and Academic Competence in Children At Risk. *Psychological Science*, *12*, 18-23.

Kazdin, A.E. (1979). Situational Specificity: The two-edged sword of behavioral assessment. *Behavioral Assessment*, 1, 57-75.

Kendler, H.H. (1987*). Historical Foundations of Modern Psychology*. Chicago, OH: The Dorsey Press.

Ladd, G., & Prolifet, W., & Muth, S. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology,* 32(6), 1008-1024.

Lewis, T.J., Sugai, G., & Colvin, G. (1998). Reducing problem behavior through a school-wide system of effective behavioral support: Investigation of a school-wide social skills training program and contextual interventions. *School Psychology Review, 27,* 446-459.

Luiselli, J.K., Putnam, R.F., Handler, M.W., & Feinberg, A.B. (2005). Whole-school positive behaviour support: Effects on student discipline problems and academic performance. *Educational Psychology*, *25*, 183-198.

Malecki, C.K., & Elliott, S.N. (2002). Children's Social Behaviors as Predictors of Academic Achievement: A Longitudinal Analysis. *School Psychology Quarterly*, 17(1), 1-23.

Marcotte, A.M., Klein, S.A., Solomon, B.G. (2010, February). *An Examination of Time Spent Teaching During Academic Learning Time.* Poster presented at the Annual Conference of the National Association of School Psychologists, Chicago, IL.

Ford, D.H. & Lerner, R.M. (1992). *Developmental Systems Theory: An Integrative Approach*. Newbury Park, CA: Sage Publications.

Mathes, P.G. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, *23*, 59-80.

McDermott, P.A., Green, L.F., Francis, J.M., & Scott, D.H., (1996). *Learning Behaviors Scale*. Philadelphia, PA: Edumetric and Clinical Science.

McIntosh, K., Chard, D.J., Boland, J.B., & Horner, R.H. (2006). Demonstration of Combined Efforts in School-Wide Academic and Behavioral Systems and Incidence of Reading and Behavior Challenges in Early Elementary Grades. *Journal of Positive Behavior Interventions, 8,* 146-154.

Menesses, K.F., & Gresham, F.M. (2010). Relative Efficacy of Reciprocal and Nonreciprocal Peer Tutoring for Students At-Risk for Academic Failure. *School Psychology Quarterly*, 24(4), 266-275

Merrell, K.W. (2008*). Behavioral, Social, and Emotional Assessment of Children and Adolescents* (3rd ed.). New York, NY: Lawrence Erlbaum Associates.

Miller, T.R. (2004). The Social Cost of Adolescent Behavior. In A. Biglan, P. A. Brennan, S. L. Foster, & H. D. Holder (Eds.). *Helping Adolescents at Risk: Prevention of Multiple Problem Behaviors*. New York, NY: The Guilford Press.

Monroe, S.M., & Harkness, K.L. (2005). Life Stress, he "Kindling" Hypothesis, and the Recurrence of Depression: Considerations From a Life Stress Perspective. *Psychological Review*, 112(2), 417-445

Najaka, S.S., Gottfredson, D.C., & Wilson, D.B. (2001). A Meta-Analytic Inquiry Into the Relationship Between Selected Risk Factors and Problem Behavior. *Prevention Science*, *2*, 257-271.

Nansel, T.R., Overpeck, M., Pilla, R.S., Ruan, J., Simons-Morton, B., & Schedit, P. (2001). Bullying Behaviors Among US Youth: Prevalence and Association With Psychosocial Adjustment. *The Journal of the American Medical Association, 25*, 2094-2100.

Nation, M., Crusto, C., Wandersman, A., Kumpfer, K.L., Seybolt, D., Morrisey-Kane, E., et al. (2003). What Works in Prevention: Principles of Effective Prevention Programs. *American Psychologist, 58*, 449-456.

Nelson, J.R., Hurley, K.D., Synhorst, L., Epstein, M.H., Stage, S., & Buckley, J.N. (2009). The Child Outcomes of a Behavior Model. *Exceptional Children, 76,* 7-30.

Northeast Foundation for Children (2003). *Responsive Classroom: Level I Workbook* (Available from the Northeast Foundation for Children, Inc, 85 Avenue A, Suite 204, Turners Falls, MA 01376).

Northeast Foundation for Children, Inc. (2009). *Responsive Classroom.* Retrieved from http://www.responsiveclassroom.org/about/aboutrc .html.

Payne, A.A., Gottfredson, D.C., & Gottfredson, G.D. (2006). School Predictors of the Intensity of Implementation of School-Based Prevention Programs: Results from a National Study. *Prevention Science, 7*, 225-237.

Pearson, P.D., & Gallagher, M.C. (1983). The Instruction of Reading Comprehension. *Contemporary Educational Psychology*, 8, 317-344.

Pianta, R.C. (1992). *The Student-Teacher Relationship Scale.* Charlottesville, VA: University of Virginia.

Pianta, R.C. (1999). *Enhancing Relationships Between Children and Teachers.* Washington, DC: American Psychological Association.

Pierce, K.M., Hamm, J.V., & Vandell, D.L. (1999). Experiences in after-school programs and children's adjustment in first-grade classrooms. *Child Development*, 70(3), 756-767.

Post, R.M. (1992). Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. *American Journal of Psychiatry*, 149, 999-1010.

Raudenbush, S.W., Byrk, A.S., Cheong, Y.F., & Congdon, R.T. (2004*). HLM6: Hierarchical Linear and Nonlinear Modeling*. Chicago: Scientific Software International.

Reschly, D.J. (1988). Special Education Reform: School Psychology Revolution. *School Psychology Review*, *17*, 459-475.

Reschly, D.J., & Hosp, J.L. (2004). State SLD Identification Policies and Practices. *Learning Disability Quarterly*, *27*, 197-213.

Rimm-Kaufman, S.E., & Chiu, Y. (2007). Promoting Social and Academic Competence in the Classroom: An Intervention Study Examining the Contribution of the Responsive Classroom Approach. *Psychology in the Schools*, *44*, 397-413.

Rimm-Kaufman, S.E., Fan, X., Chiu, Y., & You, W. (2007). The contributions of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, *45*, 401-421.

Rimm-Kaufman, S.E., & Sawyer, B.E. (2004). Primary-Grade Teachers' Self-Efficacy Beliefs, Attitudes toward Teaching, and Discipline and Teaching Practice Priorities in Relation to the "Responsive Classroom" Approach. *The Elementary School Journal*, 104(4), 321- 341.

Roberts, G., Mohammed, S.S., & Vaughn, S. (2010). Reading achievement across three language groups: Growth estimates for overall reading and reading subskills obtained with the Early Childhood Longitudinal Survey. *Journal of Educational Psychology*, 102(3), 668-686.

Ross, M.R., Powell, S.R., & Elias, M.J. (2002). New Roles for School Psychologists: Addressing the Social and Emotional Learning Needs of Students. *School Psychology Review*, 31, 43-52.

Rotter, J. (1982). Social Learning Theory. In N.T. Feather (Ed.), *Expectations and actions: Expectancy-Value models in psychology* (pp. 301-323). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sailor, W., Goetz, L., Anderson, J., Hunt, P., & Gee, K. (1988). Research on community intensive instruction as a model for building functional, generalized skills. In R. H. Horner, G. Dunlap, & R. L. Koegel (Eds.), *Generalization and maintenance: Life-Style changes in applied setting* (pp. 67-98). Baltimore: Paul H. Brookes.

Sameroff, A.J., Emde, R.N., Anders, T.F. (1989). *Relationship disturbances in early childhood: a developmental approach.* New York, NY: Basic Books.

Satcher, D. (2001). *Youth Violence: A report of the Surgeon General.* Washington, DC: U.S. Department of Health and Human Services.

Sheridan, S.M., Hungelmann, A., & Maughan, D.P. (1999). A Contextualized Framework for Social Skills Assessment, Intervention, and Generalization. *School Psychology Review*, *28*, 84-103.

Snijders, T.A.B., & Bosker, R.J. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research*, 22(3), 342-363.

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

Solomon, B. G., Klein, S., Hintze , J. M., Cressey, J., & Peller, S. (2010). *How Effective is School-Wide Positive Behavior Support in Preventing Problem Behaviors? A Meta-Analysis of the Single-Subject School-Wide Research.* Manuscript submitted for publication.

Sprague, J.R., & Walker, H.M. (2005). *Safe and Healthy Schools: Practical Prevention Strategies.* New York, NY: The Guilford Press.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-407.

Sugai, G., Horner, R.H., Dunlap, G., Hieneman, M., Lewis, T.J., Nelson, C.M., et al. (2000). Applying Positive Behavior Support and Functional Behavioral Assessment in Schools. *Journal of Positive Behavior Support*, *2*, 131-143.

Sugai, G., Horner, R.H., & Gresham, F.M. (2002). Behaviorally Effective School Environments. In M. R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 373-401). Bethesda: National Association of School Psychologists.

Teo, A., Carlson, E., Mathieu, P.J., Egeland, B., & Sroufe, L.A. (1996). A Prospective Longitudinal Study of Psychosocial Predictors of Achievement. *Journal of School Psychology*, *34*, 285-306.

Touchstone Applied Science Associates (2002). *Degrees of Reading Power program*. Brewster, NY: Author.

van Lier, P. A. C., Muthen, B. O., van der Sar, R. M., & Crijne, A. A. M. (2004). Preventing disruptive behavior in elementary schoolchildren: Impact of a universal classroom-based intervention. *Journal of Consulting and Clinical Psychology*, 72, 467-478.

Vannest, K.J., & Parker, R.I. (2010). Measuring Time: The Stability of Special Education Teacher Time Use. *The Journal of Special Education*, 44(2), 94-106.

Vygotsky, L.S. (1978). *Mind in Society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walker, H.M., & Severson, H. (1990). *Systematic Screening for Behavior Disorders (SSBD)*. Longmont, CA: Sopris West.

Walker, H.M., Severson, H., & Feil, E.G. (1995). *Early Screening Project (ESP).* Longmont, CA: Sopris West.

Weisz, J.R., Sandler, I.N., Durlak, J.A., & Anton, B.S. (2005). Promoting and Protecting Youth Mental Health Through Evidence-Based Prevention and Treatment. *American Psychologist, 60,* 628-648.

Wentzel, K.R. (1993). Does Being Good Make the Grade? Social Behavior and Academic Competence in Middle School. *Journal of Educational Psychology, 85*, 357-364.

Wentzel, K.R., & Watkins, D.E. (2002). Peer Relationships and Collaborative Learning as Contexts for Academic Enablers. *School Psychology Review, 31*, 366-377.

Wilson, D.B., Gottfredson, D.C., & Najaka, S.S. (2001). School-Based Prevention of Problem Behaviors: A Meta-Analysis. *Journal of Quantitative Criminology, 17*, 247-272.

Woodcock, R.W. (1998). *Woodcock Reading Mastery Tests-Revised*. Circle Pines, MN: American Guidance Service (AGS).

Ysselydyke, J., Burns, M., Dawson, P., Kelley, B., Morrison, D., Ortiz, S., Rosenfield, S., & Telzrow, C. (2006). School Psychology: blueprint for training and practice III. Retrieved from the National Association of School Psychologist's website: http://www.nasponline.org/resources/blueprint/index.aspx.

Zill, N., & Nord, C.W. (1994). *Running in Place: How American Families are Faring in a Changing Economy and an Individualistic Society*. Washington, DC: Child Trends, Inc.

Zirkel, P.A., & Krohn, N. (2008). RTI After IDEA. *Teaching Exceptional Children, 40*, 71-73.