A COMPARISON OF TEACHER AND STUDENT RESPONSIVENESS FOR MALE STUDENTS AT HIGH AND LOW RISK FOR EXTERNALIZING BEHAVIOR PROBLEMS

By

Daniel Montagne Maggin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

August, 2009

Nashville, Tennessee

Approved:

Professor Joseph H. Wehby

Professor Donald L. Compton

Professor David S. Cordray

Professor Lynn S. Fuchs

Professor Kathleen L. Lane

To my incredible and supportive wife Wendy and beautiful daughter Cora. Thank you for your patience and encouragement.

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of the Experimental Education Research Training (ExpERT) pre-doctoral fellowship provided by the Institute for Educational Sciences (IES) and the Vanderbilt Behavior Research Center (VBRC). I am especially indebted to Dr. Joseph H. Wehby, the chairman of my Dissertation Committee, my advisor, and friend who has been supportive of my career goals and who worked actively to provide me with the resources to pursue my degree and the present research. As my mentor, he has taught me about professionalism, good science, and always finding the "hook." His example will always help to guide my life both personal and professional.

I am grateful to all those with whom I have had the pleasure to work during this and other related projects. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and pursuing careers in education research. I would especially like to thank Mrs. Terrell McGuire for her help in organizing logistical concerns related to my Dissertation project. She has been an invaluable team member and driving force behind completion of this project.

Finally, I would like to thank my family who has always supported me in pursuing my personal and professional goals. I would like to thank my parents, whose love and guidance are with me in whatever I pursue. They are the ultimate role models. Most importantly, I wish to thank my loving and supportive wife, Wendy, and my incredible daughter, Cora, for her unending love and inspiration.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

CLASSROOM INTERACTIONS BETWEEN TEACHERS AND STUDENTS AT RISK FOR EXTERNALIZING BEHAVIOR DISORDERS

The academic and behavioral difficulties exhibited by students at risk for developing externalizing behavior disorders (EBD) provide educators with unique challenges. Students at risk for EBD tend to display aggression, noncompliance, and other conduct problems at higher rates and with more stability than typically developing peers (Kauffman, 2005; Lane, 1999; Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998). For example, students at risk for EBD have been observed to emit more negative statements toward teachers (Moore & Simpson, 1983) and to display more inappropriate behaviors (Landrum, Tankersley, & Kauffman, 2003; Walker, Shinn, O'Neill, & Ramsey, 1987) than typically developing peers. In fact, students at risk for EBD can be distinguished from the general population at an early age for their behavioral excesses (Stormont, 2002).

In addition to behavioral issues, research has indicated that students at risk for EBD also confront significant academic deficits (Deater-Deckard, Dodge, Bates, & Pettit, 1998). A recent meta-analysis found that students with EBD read nearly three-quarters of a standard deviation below their same aged peers (Reid, Gonzalez, Nordness, Trout, & Epstein, 2004). Such academic deficits are often proposed as a contributing factor to the development of EBD in elementary aged students (e.g., Hinshaw, 1992). The academic difficulties displayed by students with and at risk for EBD are typically more pronounced than those of students with learning disabilities (LD) despite a similarity in intelligence levels (Sabornie, Cullinan, Osborne, & Brock, 2005). In fact, some evidence has indicated that behavioral issues might

impede academic progress more than learning difficulties over time (Anderson, Kutash, & Duchnowski, 2001).

Coercive Interactions

The academic and behavioral deficits exhibited by students at risk for EBD make educating them particularly challenging. In fact, the strain of instructing and managing students with EBD often place teachers and students at risk for developing coercive interactions. In its truest form, coercion theory (Patterson, 1982) is a formulation of the causes of conduct problems in children. According to coercion theory, the development of antisocial behavior begins with mild disruptive behaviors that evolve into more severe problems such as aggression. The hypothesized mechanism in this evolution from mild to severe behaviors is the micro-social interactions between child and adult (Patterson, 1976). Specifically, coercive interactions are driven by a cycle of negative reinforcement in which child and adult behavior persists because it results in the termination of aversive behavior from the dyadic partner. For example, a student might comply with repeated requests by the teacher to turn in his homework only to make the successive requests stop. Conversely, a teacher might respond to a student tantrum by withdrawing a particular demand, such appeasement is negatively reinforcing to the teacher because it removes the unpleasantness of the tantrum. Therefore, the teacher will be more likel to withdraw task demands in the future. Patterson describes this type of interaction as a coercion trap. Both the teacher and student are participating in a system of interactions that has short-term benefits (i.e., removal of aversive behavior), but establishes a framework for long-term behavior problems.

Investigations comparing patterns of classroom responding by teachers and students with and without EBD have provided preliminary evidence for the presence of coercive interactions. For instance, teachers have been shown to provide less instruction and positive attention to students with antisocial behaviors (e.g., Carr, Taylor, & Robinson, 1991; Montague & Rinaldi, 2001; Taylor & Romanczyk, 1994; Wehby, Symons, Canale, & Go, 1998). Within a framework of coercion, such modifications of teacher behavior occur to reduce the frequency and intensity of disruptive behaviors (e.g., Gunter, Denny, Jack, Shores, & Nelson, 1993; Gunter, Shores, Jack, Denny, & DePaepe, 1994). Additional descriptive research has lent support to the notion that negative reinforcement can have deleterious effects on the academic abilities of students (McIntosh, Horner, Chard, & Dickey, 2008). McIntosh and colleagues found that students at risk for escape-maintained problem behaviors performed substantially worse than students whose behavioral problems were maintained by a different function. The authors couched their conclusions within a framework of coercion and determined that its likely students with escape-maintained problem behaviors typically receive less academic instruction and fewer academic demands than students with other types of behavior problems.

Although some empirical support has buttressed the notion that coercion is present within interactions between teachers and students at risk for EBD, little is currently known about the specific processes that occur to regulate student prosocial and antisocial behavior during instructional periods. Better awareness of these micro-social processes can potentially lead to stronger interventions and a more complete understanding of the classroom behavior of both teachers and students. The present study represents an attempt to address issues regarding the conceptualization and measurement of micro-social classroom interactions between

individual students and their teachers. Specifically, the notion of and empirical justification for considering the micro-social process of *responsiveness* between teachers and students is presented. According to Patterson (1982), responsiveness is both an indicator of and contributor to the development of coercive interactions. As an indicator, the timely responding and shared content between two people are markers for non-coercive relationships. Likewise, if responding is not timely and content is not shared, coerciveness might be present. As a contributor, the lack of responsiveness between two people over time might lead to an increase in coercion within their interactions.

The following sections will describe the school-based literature on teacher-student responsiveness (and other proxy measures) with three goals in mind. The first goal is to justify the importance of considering teacher-student responsiveness in classroom-based interactional research. A second goal is to critique traditional approaches to measuring teacher-student interactions and responsiveness and to describe a novel approach using direct observation techniques. The final goal is to describe the purpose, contributions, and guiding research questions of the study.

Responsiveness

The extent to which an individual's response reciprocates another's behavior within a timely fashion is referred to as responsiveness (e.g., Harrist & Waugh, 2002; Patterson, 1982). A teacher with high responsiveness would provide praise, reprimands, instruction, and penalties that match student behavior. Responsive interactions are important because they continually establish the reinforcement and punishment contingencies of the classroom (Herring & Wahler, 2003). In other words, students can expect that misbehavior will be met

with a consequence or that an academic question will be answered if interactions are characterized by responsiveness. Although the majority of research on responsiveness has pertained to parent-child interactions, there is some indication that the degree of teacher responsiveness depends on the perceived quality of student behavior.

Skinner and Belmont (1993), for example, examined the association between teacher behavior and levels of student engagement during instruction. Results indicated that students with higher levels of academic engagement perceived their teachers as responding more contingently (i.e., with responsiveness) to the classroom behavior of students. A notable limitation of the Skinner and Belmont study was the use of teacher and student report of classroom behavior (e.g., student engagement, teacher attention; perceived contingency). Van Acker, Grant, and Henry (1996) addressed this limitation by using direct observation to compare teacher response patterns toward students with moderate and high aggression levels during instructional activities. Results demonstrated that teacher behavior toward students was dependent on whether the student was a moderate aggressor or high aggressor. Although both high and low aggressors were shown to have a similar rates of correctly answered academic questions, teachers were more likely to direct questions toward low aggressors and praise low aggressors for correct answers. In contrast, students at high risk for aggression were more likely to receive a reprimand for inappropriate behavior. Results from these studies indicate that teachers were less responsive to the appropriate behavior of students with greater behavioral issues and more responsive to inappropriate behavior. Such an imbalance in responsiveness toward students with the greatest behavioral needs may result in the (a) extinction of appropriate behavior by ignoring its occurrence and (b) promotion of undesired behavior through ineffective punishment strategies.

Additional evidence supports the notion that teachers respond differently to students with varying behavioral profiles and establish contingencies that promote inappropriate behavior. Nelson and Roberts (2000) compared sequences of teacher-student interactions following an inappropriate behavior for disruptive and non-disruptive students enrolled in general education classrooms. Results indicated that both groups had similar instances of non-compliant behavior. However, disruptive students were more likely to receive a reprimand following inappropriate behavior, while non-disruptive students were more likely to receive a redirection. Though students in both groups demonstrated similar levels of disruptiveness, students considered to be disruptive received a disproportionate amount of negative attention. Coupled with data that suggests teachers infrequently use positive methods to reinforce appropriate behavior (e.g., Wehby, Symons, & Shores, 1995; Shores, Jack, Gunter, Ellis, DeBriere, & Wehby, 1993), these findings lend credence to the notion that teacher behavior might encourage inappropriate student responding.

The findings presented across these studies seem to suggest that teacher responsiveness may depend on both the perceived level of student antisocial behavior and their actual classroom behavior. For example, teachers have been shown to modify their behavior, to varying degrees, toward more challenging students (e.g., Carr et al., 1991; Montague & Rinaldi, 2001; Nelson & Roberts, 2000; Skinner & Belmont, 1993; Taylor & Romanczyk, 1994; Van Acker et al., 1996; Wehby et al., 1998). Teachers have been shown to be more likely to respond to inappropriate rather than appropriate student behavior (Nelson & Roberts, 2000; Shores et al., 1993; Wehby et al., 1995). According to coercion theory, these teacher-student interaction patterns establish classroom contingencies that might increase the probability of a student emitting a negative behavior. Specifically, teachers seemingly fail to

reinforce appropriate behavior. Furthermore, misbehavior may ultimately be reinforced through ineffective punishment techniques such as yelling, threatening, and other forms of ineffective punishment (Gunter et al., 1994; Harrison, Gunter, Reed, & Lee, 1996; Patterson, 1982). Ensuring that teacher responses are aligned with student behavior might increase the likelihood that students with EBD react in accordance with classroom norms. In fact, Herring and Wahler (2003) reported that teacher responsiveness, defined as the extent to which a teacher's response matches student behavior, protected against problematic classroom behaviors in a sample of typically developing students.

Measurement Issues

The present study is an attempt to extend the literature on teacher-student interactions by comparing levels of teacher and student responsiveness associated with students at high and low risk for EBD. A primary challenge was to develop a valid and reliable measurement system for assessing responsiveness. Past measures of dyadic responsiveness have been plagued by two measurement issues. First, operational definitions of responsiveness have lacked consistency and precision (Lindsey, Colwell, Frabutt, Chambers, & MacKinnon-Lewis, 2008). The tendency of researchers to group several constructs into a single characterization of responsiveness has limited their ability to draw inferences and generalizations both within and across studies. For instance, Lindsey and colleagues compared findings on three subconstructs traditionally used in the literature to define responsiveness. These subconstructs were (a) the degree of coordinated and contingent behavioral and verbal exchanges, (b) shared affect, and (c) balanced conversational participation. Findings revealed that, in fact, each was a distinctly independent construct. This suggests that using a constellation approach to measuring

responsiveness might be concealing important relationships involving the subordinate factors. Therefore, the process of constructing a definition of responsiveness that allows for clear measurement and unambiguous conclusions to be drawn was undertaken.

A second concern with past measurement systems is the failure to capture the sequential nature of responsiveness. Despite conceptually recognizing the sequential dependencies of dyadic responses, measurement approaches have failed to account for such patterning in data collection (e.g., Criss, Shaw, Ingoldsby, 2003; Deater-Deckard, Atzaba-Poria, & Pike, 2004) or data analysis (e.g., Isabella & Belsky, 1991; Rocissano & Yatchmink, 1983). This limitation is due to an over-reliance on Likert scales to measure the degree of responsiveness between older children and their caregivers (e.g., Criss et al., 2003; Deater-Deckard et al., 2004; Lindsey et al., 2008). The advantage of Likert systems is that the proficiency of observed behaviors can be rated on a global continuum. However, the use of Likert scales limits the ability of researchers to measure the interactional sequences that are central to the notion of responsiveness.

In order to address the concerns with past attempts to measure responsiveness, the following considerations were made. First, operational definitions of responsiveness were differentiated from other subconstructs (e.g., affect; balanced participation). A focused definition was developed to allow for more accurate measurement procedures and stronger inferences from study results. The focus of the operational definitions used for the present study was on alignment of content between a teacher-student pair and timeliness of response. Second, direct observation was used to assess the degree of responsiveness present in teacher and student interactions. Specifically, sequential analysis (Bakeman &Gottman, 1997) was used to maintain the sequential nature of the responsiveness variables. Sequential analysis is a

set of data collection and analysis techniques that allows for an estimate of the probability that a given behavior increases or decreases the occurrence of subsequent behaviors (Bakeman & Gottman; Yoder & Feurer, 2000).

Purpose and Contribution of Study

The purpose of this descriptive study was to extend the micro-social understanding of teacher-student interactions during periods of academic instruction. Specifically, direct observation was used to (a) quantify summary level data of naturally occurring rates of both positive and negative behaviors for teachers and students during academic instruction and (b) conduct a series of sequential analyses on the responsiveness of teachers and students. Following the quantification of summary-level data and computations of sequential associations, group comparisons were conducted to determine if teachers responded differently to students at high and low risk for developing externalizing behavior problems. In addition, measures of student responsiveness were computed and compared across high and low risk groups to determine if students responded to teacher behaviors differently based on risk status. The following research questions helped to guide the study:

1. Is there a difference in the absolute amount of positive, negative, or instructional behavior directed from teachers toward students at high and low risk for developing EBD?
2. Is there a difference in the absolute amount of positive, negative, or academic behavior emitted by students at high and low risk for developing EBD?

3. Are teachers more or less responsive toward students at low risk for developing EBD than students at high risk for developing behavior problems during academic interactions?
4. Are teachers more or less responsive toward students at low risk for developing EBD than students at high risk for developing behavior problems during conduct-based interactions?
5. Are students at low risk for developing externalizing behavior problems more or less responsive to teacher instructional behaviors than students at high risk for developing EBD?
6. Are students at low risk for developing externalizing behavior problems more or less responsive to teacher management cues than students at high risk for developing behavior problems?

CHAPTER II

METHOD

Participants

A total of 25 general education classrooms were sampled for this study. All classrooms served students in 1st through 3rd grades. Two teacher-student dyads were sampled from each classroom. The dyads consisted of the classroom teacher paired with (1) a student at high risk for externalizing behavior problems and (2) a student at low risk for externalizing behavior problems. Past research on teacher-student dyads has indicated differences in teacher and student behavior based on gender (e.g., Baker, 1999; Skiba, Poloni-Staudinger, Gallini, Simmons, & Feggins-Azziz, 2006). Therefore, only male students were recruited for this study in order to control for potential gender differences. In addition, students with an individualized education plan (IEP) or behavior intervention plan (BIP) were excluded from participation due to concerns of specialized instruction or management plans that might impact teacher or student behavior.

Procedures

Participant selection and screening. All classrooms were drawn from the same urban, southeastern school district. The district serves approximately 75,000 students enrolled across 133 schools (pre-kindergarten to high school). As of the 2008-2009 school year, the district reports its ethnic makeup to be 33.14% Caucasian, 48.02% African-America, 14.99% Hispanic, 3.36% Asian, and 0.26% students to be of other backgrounds. In addition, 72.80%

of students in the district were reported to be economically disadvantaged with approximately 14.10% of all students having qualified for special education services. According to district reports on the achievement of students in kindergarten to eighth grade, 85.00% of students tested proficient or advanced in reading on the state test and 88.00% tested proficient or advanced in mathematics. These scores are from the 2007-2008 school year.

To be eligible to participate, each classroom had to be housed in a school in which (a) 75% of the student population received free or reduced lunch and (b) less than 50% of students were proficient on the state reading exam. Following initial screening of area schools, an email was sent to principals requesting permission to address faculty members about participating in the study. A total of 22 principals were contacted with 8 (36.36%) expressing interest in participating. One school had to be dropped due to an inability to arrange a time to address the faculty. Following permission to address faculty members, teachers were addressed by either the first or second author during regularly scheduled faculty meetings. The parameters of the study (i.e., teacher and student rights and responsibilities) were outlined during these meetings. Teachers that were already participating in other research projects were declared ineligible to participate in this study. Across each of the seven schools, a total of 66 teachers were eligible to participate. Of these, 41 (62.12%) agreed to participate. An additional 16 teachers had to be dropped either because there were no eligible students in the classroom or qualifying students did not return consent forms. Therefore, of the 66 eligible teachers 25 (37.88%) consented for participation and had eligible students in their classroom. Of the 25 participating classrooms, 8 were drawn from 1st-grade, 8 were drawn from 2nd-grade, and 9 were drawn from 3rd-grade.

Teacher Participants. Of the 25 participating teachers, 23 were female (91.67%). In addition, 21 of the teachers were Caucasian (84.00%) with the remaining 4 teachers being African-American (16.00%). Teacher participants had an average of 15.60 (*SD* = 11.53) years of experience with a range of 1 to 35 years. Nearly half of the participating teachers held a Master's degree (44.00%). A fifth of the teachers (20.00%) had earned credits over their Master's degree with one of these teachers having earned a doctorate and four others having earned credits above their Master's. Of the remaining teachers, three held a Bachelor's degree in education (12.00%) and six had earned credits beyond their Bachelor's degree (24.00%). More than three quarters of the reported degrees were in general education or elementary education (76.00%). The remaining degrees were in early childhood education (8.00%), administration (8.00%) reading (4.00%), and counseling (4.00%).

All teachers completed a researcher developed measure, *The Class Characteristics and Language Arts Survey*, used to assess the overall classroom characteristics and language arts curriculum. The survey had a total of 23 items across 3 sections. The first section focused on personal characteristics of the teacher including gender; years of experience; grade taught; teacher ethnicity; and education level. The second section assessed characteristics of the learning environment including the total number of students in the class; demographic features of the students; and perceptions of the academic and behavioral functioning of students. The third section focused on the language arts curriculum. Since a whole-group language arts instruction was targeted for observations, it was important to assess similarities and differences of the curriculum across classrooms. Items for this portion of the survey focused on the use of a commercial reading program; the homogeneity of reading abilities of the

students present during language arts; and the frequency of particular activities used by the teacher.

Participating Classrooms. As to be expected, demographic variation between classrooms was observed. For instance, the average number of students per classrooms was approximately 17.33 (*SD* = 2.44) with a range of 13 to 24 students. Few students within the sampled classrooms received special education services (6.42%) with 16 of the classrooms having 1 or less students with an IEP. The highest reported total of students receiving special education services was five. The ethnic makeup of these classrooms approximated figures reported by the school district. Most students in the sampled classrooms were African-America (50.00%) or Caucasian (30.00%). Although a moderate number of Hispanic students were represented across the classrooms (15.00%), a majority of these students were enrolled in three classrooms within the same school. Few students of Asian descent (2.52%) or from other backgrounds (2.98%) were represented in the participating classrooms. Similar to that of the Hispanic population, these students tended to be grouped together within a particular classroom or school.

In addition to classroom demographic data, information on the reading and behavioral functioning of each classroom was collected using teacher report. Data for this were collected using the researcher developed survey designed to assess the language arts curriculum and overall classroom characteristics. Considerable variability was found between classrooms on reported levels of reading and behavioral functioning. For example, approximately one quarter of the students in the sampled classrooms were reported to be reading above grade level (26.37%). The range of reported percentages for those students reading above grade level was from 0% to 75%. Less than half of the students in participating classrooms were reported to be

reading at grade level (45.32%). The range of reported percentages for students reading at grade level was between 10% and 80%. More than one quarter of the students in the sampled classrooms were reported to be reading below grade level (28.32%). The range of reported percentages for students reading below grade level was from 0% to 70%.

Similar variability was observed from teacher reports of the behavioral functioning of students within the sampled classrooms. Less than half of the students across these classrooms were reported to have no behavior problems (47.66%). However, the range of reported percentages for students having no behavior problems was from 5.00% to 95.00%. More than a quarter of the students in the participating classrooms were reported to have moderate behavior problems (29.13%). Moderate behavior problems were defined as acting out once or twice a week during instructional periods. The range of students reported to have moderate behavior problems was between 5.00% and 75.00%. Teachers reported that nearly a quarter of students in the sampled classrooms had persistent problem behaviors (23.59%). Persistent problem behaviors were defined as acting out more than twice per week during instructional periods. Again, variability between classrooms was high with a range of 0% to 60% of students reported to have persistent behavioral problems.

Student Participants. The student participant process began with teachers sending a Phase 1 consent form home with all male students in the class. This consent form requested permission to screen male students into the study. To identify students at high and low risk for EBD, the teacher completed the *Student Risk Screening Scale* (Drummond, 1994) for those students who were given permission to participate. The SRSS (Drummond) is a low-cost, seven-item measure used to conduct mass screenings of elementary aged students who are at-risk for developing externalizing behavior problems (Drummond, Eddy, & Reid, 1998a,

1998b). To complete the SRSS, teachers rate each student by using a 4-point Likert scale (*never* = 0, *occasionally* = 1, *sometimes* = 2, *frequently* = 3) on seven items: (1) steals; (2) lies, cheats, sneaks; (3) behavior problems; (4) peer rejection; (5) low achievement; (6) negative attitude; and (7) aggressive behavior. Scores range from 0 to 21, with higher scores indicating more risk for externalizing behavior problems. Risk status ranges from low (0 to 3) to moderate (4 to 8) to high (9 to 21). The SRSS has been shown to successfully distinguish between young students with and without behavior problems. In addition, the SRSS significantly correlates (r = .79) with the aggression subscale of the more intensive Child Behavior Checklist (CBCL; Walker, Ramsey, & Gresham, 2004). Further evidence of the utility of the SRSS in differentiating between children at high and low risk for externalizing behavioral problems is found in studies of the tools' predictive validity in which SRSS scores at 18-months were related to negative behavioral and academic outcomes at 10 years old (Drummond, Eddy, Reid, & Bank, 1994). The SRSS is an empirically validated screening tool (Drummond, 1998a, 1998b) that classifies students as being at low, moderate, and high risk for developing externalizing behavior problems. Due to concerns about the ability of the SRSS to differentiate between adjacent categories (e.g., low and moderate risk; moderate and high risk) (see Lane, Little, Casey, Lambert, Wehby, Weisenbach et al., 2008), only students from the low and high risk categories were deemed eligible. If multiple students in a given classroom were screened into the same risk category students were randomly selected for participation. Random selection occurred by placing student names on slips of paper and putting those slips into a container. A research assistant would then select a slip of paper from the container and the student whose name was on the paper was chosen to participate.

Following the screening procedure, 50 students, 2 from each classroom, were enrolled in the study. Comparisons between students at high and low risk for EBD are presented in Table 1. The average age of participating students across the whole sample was 7.71 years (*SD* = 1.01) with a range of 6 to 10 years. There was no statistical difference between the two groups in terms of age F (1, 49) = .73, p = 40 with students at high risk (M = 7.83, SD = 1.09) being, on average, slightly older than students at low risk (M = 7.58, SD = .93). In addition, there were no group differences in ethnicity χ^2 (3, N = 50) = 3.86, p = .28. Analyses demonstrated that students did, in fact, differ on teacher reports of externalizing behaviors. The SRSS scores showed that high risk students had higher scores (M = 13.68, SD = 3.38) than low risk students (M = 2.12, SD = 1.69). This difference was shown to be statistically significant F (1, 49) = 234.39, $p < .001$ with a one-way ANOVA.

Table 1.

Student Participant Demographic Data by Risk Category

Variable	*Level*	High Risk				Low Risk			
		M	SD	n	%	M	SD	n	%
Ethnicity	African-American			14	56.00			12	48.00
	Asian			0	0.00			1	4.00
	Caucasian			8	32.00			5	20.00
	Hispanic			3	12.00			7	28.00
WJ-III (SS)	Broad Reading	93.88	16.48			99.20	18.21		
	Basic Reading	100.80	15.20			105.04	14.10		
	Letter-Word ID	99.52	14.61			103.16	14.92		
	Reading Fluency	94.10	13.25			100.33	12.57		
	Passage Comprehension	92.84	15.37			97.36	12.55		
	Word Attack	101.68	14.36			106.44	11.76		
ORF (RS)		63.92	40.17			77.18	40.31		
SRSS (RS)		13.68	3.38			2.12	1.69		
TRF (SS)	Externalizing*	69.13	7.80			48.58	7.16		
	Attention	50.04	0.20			50.00	0.00		
	Rule Break*	66.54	6.20			53.08	3.56		
	Aggressive*	70.38	10.70			52.42	3.88		
	Oppositional*	66.96	5.12			51.50	2.84		
	Conduct*	66.04	9.28			51.83	3.60		

* indicates a *p*-value < .05 between risk categories

In order to verify results from the SRSS that students in the high and low risk groups differed on levels of externalizing behavior problems, teachers also completed the externalizing subscale of the *Teacher Report Form* (TRF). The TRF (Achenbach & Rescorla, 2001; r = .89) is a subsection of the Achenbach System of Empirically Based Assessment (ASEBA) which is comprised of an integrated set of forms for assessing student competencies, adaptive functioning, and problems. The current edition of the TRF is normed for students between the ages of 6 and 18. For the purposes of the present study, only those items from the externalizing subscale of the TRF were used. Therefore, teachers completed a revised rating scale of 32 items. Example items include: (1) argues a lot; (2) defiant, talks back to staff; and (3) Cruelty, bullying, or meanness to others. Each item is rated on a 3-point Likert scale ranging from 1 to 3 (1 = *not true*, 2 = *somewhat or sometimes true*, 3 = *very true or often true*). Students with externalizing scores that are above the 90th percentile are considered to be within a clinical range of behavioral problems. Students whose externalizing scores are between the 84th and 90th percentiles are considered to be within the borderline clinical range. The TRF is a psychometrically sounds instrument that is widely regarded as the gold standard for the study of behavioral problems in students (Walker & Severson, 2002).Analyses comparing the high and low risk groups on TRF scores showed that the high risk students (M = 69.13, SD = 7.80) were rated by their teachers as having more behavior problems than the low risk students (M = 48.58, SD = 7.16) F (1, 49) = 90.33, p < .001. Further inspection of Table 1 shows that students in the high risk group also had statistically significant differences on four of the five syndrome scales that comprise the broader externalizing subscale of the TRF. The only syndrome scale on which high and low risk students did not differ was attention F (1, 49) = 1.00, p = .32.

Following inspection of TRF scores to ensure that students in the high and low risk categories differed on levels of externalizing behavior problems, a Phase 2 consent form was sent home with both the high and low risk students. The Phase 2 consent form requested permission from the parents to observe their student during whole-class instruction. Students that returned the Phase 2 consent form indicating parent approval were then observed during whole-class instructional periods (see observational methods section) and given a brief reading assessment. The reading assessment consisted of four subscales from the Woodcock-Johnson-III Test of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001). The WJ-III provides a set of standardized achievement tests that assess a range of academic competencies. The battery as a whole provides information on (a) oral language (i.e., oral language, oral expression, and listening comprehension); (b) reading (broad reading, basic reading skills, and reading comprehension); (c) mathematics (broad math, math calculation skills, and math reasoning); and (d) written language (broad written language, basic written skills, and written expression). Based on a nationally representative sample, the WJ-III achievement battery has strong psychometric properties with reliability estimates of .80 and higher. For this study, the broad reading cluster scores were used to identify differences in the reading ability of students at high and low risk for externalizing behavior problems. Reading ability was identified as an important variable because language arts periods were targeted for direct observations. The broad reading score is comprised of four subtests including letter-word identification, reading fluency, passage comprehension, and word attack. Internal consistency reliabilities exceed .90 for all of the subtests used indicating excellent psychometric properties. Although students at low risk (M = 99.20, SD = 18.21) tended to score higher on the broad reading scale than students at high risk (M = 93.88, SD = 16.48) these differences were not statistically

significant F (1, 48) = 1.17, p = .28. None of the scores derived from the WJ-III were found to be significant (all p's > .10).

In addition to the WJ-III, students were given a timed Oral Reading Fluency measure (ORF; National Behavior Research Coordinating Center, 2004). The ORF is a curriculum-based measure in which the student is given two passages to read out loud for one-minute each. Both passages are written for a first grade level. As the student reads out loud, a research assistant completed a running record with all correctly read words being tallied to produce a score. For the present study, the student's total number of correctly read words was averaged across the two passages. This averaged score was used to compare the reading ability of students at high and low risk for EBD. Findings from the ORF measure were similar to those found with the WJ-III. Specifically, low risk students (M = 77.18, SD = 40.31) tended to score higher on the measure than high risk students (M = 63.92; SD = 40.17) but these differences were not found to be significant F (1, 49) = 1.33, p = .25.

Observational Methods

Development and validity of coding system. An observational coding scheme requires the sampling of observable behaviors that relate to the construct being measured (Bakeman & Gottman, 1997). Since the research questions relate to teacher and student responsiveness during instructional periods, those behaviors emitted by teachers and students during academic instruction were of the greatest interest. In addition, the primary outcome variable was a measure of sequential association between teacher and student behavior and required codes to be mutually exclusive (Yoder & Feurer, 2000). That is, each code could only be associated with a particular event. Three steps were undertaken to ensure that the codes and operational

definitions were mutually exclusive and comprehensive of teacher and student behaviors during whole-class instructional periods.

First, the literature on adult-child responsiveness was reviewed to assist with identifying those constructs that could be operationally defined. Findings from this literature review revealed that traditional measures of responsiveness tended to group several constructs together including shared affect, shared content, and balanced participation. This blending of subconstructs into a single definition of responsiveness has limited the ability to draw firm conclusions about the most salient construct related to outcome variables. In addition, such grouping limits the data analysis strategies to methods that are not able to reflect the sequential nature of dyadic responsiveness. Therefore, the operational definition of teacher and student responsiveness for the present study focused on the most salient aspects of the construct, namely, shared content and timeliness of response. Such an emphasis led to the development of the following operational definitions: (a) *the extent to which a teacher's response is mutually focused and reciprocates a student's behavior within three-seconds from the initiation of student behavior* and (b) *the extent to which a student's response is mutually focused and reciprocates a teacher's behavior within three-seconds from the termination of teacher behavior.*

The second step undertaken was a review of the literature to identify those classroom behaviors that are commonly sampled in studies of classroom interactions between teachers and students. Broadly, the categories of behavior typically coded could be classified as being academically-based (e.g., instruction; hand-raising), conduct-based (e.g., reprimand; disruptive behavior), or socially-based (e.g., talking with other students about social topics; engaging a peer) (Brophy & Good, 1986). For the present study, social behaviors (e.g., unapproved

whispering to neighbor) were collapsed into the conduct-based behavior category due to the focus on the relevancy of teacher and student responses. In other words, for responsive interactions to be accurately measured behaviors must be able to be considered as "appropriately aligned" or "inappropriately aligned" with partner behavior. The alignment of social behaviors is more difficult to determine from observational codes than academic-based or conduct-based behaviors. Therefore, the superordinate categories for the observational system were academically-based behaviors and conduct-based behaviors. Student and teacher behaviors were classified directly into these superordinate categories. In some cases, behaviors could be considered both an academically and conduct based. Such overlap was not deemed a concern since there is not an infringement on computing sequential associations between behaviors. Table 2 lists the behavior codes, operational definitions, and superordinate categories used for data analysis.

Table 2.

Domains, Operational Definitions, and Data Analysis Categories for Interaction Codes

Domain	*Code*	*Operational Definition*	*Category for Data Analysis*
Teacher Behavior	Instructional Talk	Verbal statements relating to instructional tasks directed towards the target student or to the group of which the target student is a part.	Academic Behavior
	Academic Prompt	Any verbal or physical statements that request a verbal or written response related to academic material from the target student or to the group of which the target student is a part.	Academic Behavior
	Acknowledgement	Any verbal or physical statement that indicates the teacher is formally recognizing a student.	Academic Behavior Conduct Behavior
	Academic Feedback	Statements that provide information and / or clarification of an academic concept.	Academic Behavior
	Command	Statements that require specific and immediate physical responses by the target student or group of students of which the target student is a part.	Conduct Behavior
	Reprimand	Statements or gestures made by the teacher to the target student as an individual or group indicating disapproval of the student's social behavior.	Conduct Behavior
	Praise	Statement of gesture that indicates approval of behavior over and above an evaluation of adequacy or acknowledgement to either a social behavior or academic response.	Academic Behavior Conduct Behavior
Student Behavior	Request for Attention	Any request for attention that utilizes appropriate, pro-social methods (e.g., hand raise).	Academic Behavior Conduct Behavior

Table 2 (continued).

Academic Engagement	Student is engaged on assigned or approved activity. Engagement is signaled by looking, writing, and waiting in a manner that meets the expectations of the activity. Student is not engaged if he does not attend to task for more than 5-seconds.	Academic Behavior
Compliant Response	Appropriate physical and verbal behaviors that are initiated following a teacher mand (e.g., academic prompt; command).	Academic Behavior Conduct Behavior
Disruptive Behavior	Inappropriate physical and verbal behaviors that violate classroom norms. Range from minor behaviors (e.g., call-outs) to severe behaviors (e.g., aggression).	Conduct Behavior

The third step in the development of the coding scheme was to examine its content validity. This was achieved by recruiting eight experts familiar with conducting quantitative observational research in elementary school classrooms to complete three activities designed to test the schemes' comprehensiveness and mutual exclusivity. The eight experts were drawn from the fields of special education (n = 4), psychology (n = 3), and curriculum and instruction (n = 1). Each expert was contacted via email to participate and, upon consent, was mailed a packet of materials. The materials consisted of (a) the proposed coding manual with each code and operational definitions, (b) a sheet of paper with the codes without operational definitions, and (c) a content validity protocol. Experts were asked to complete the study in a timely fashion and to return all materials back to the lead author.

The first task experts were asked to complete focused on ensuring that operational definitions for each code were clear. Specifically, participants were to match a set of 30 behaviors (e.g., call-out, talking to neighbor during instruction) to behavioral domains and codes included in the coding scheme (e.g., teacher instruction). One of the items had to be removed from consideration due to a misprint in the content validity protocol. Of the remaining 29 items, 23 (79.31%) had exact agreement across all participants. Five (17.24%) of the remaining 29 total items had one disagreement. One (0.03%) of the 29 total items had two disagreements. The lowest level of agreement across experts on the 29 eligible items was, therefore, 75.00% with this occurring only once. Such findings indicate that the operational definitions for each code did, in fact, conform to the behaviors that they purported to measure.

The second activity was an examination of the mutual exclusivity of codes. In order to determine if codes were mutually exclusive, an adapted version of the item-objective congruence index (Rovinelli & Hambleton, 1977) was developed. Within test theory, the item-

objective congruence index is used to assess the degree to which a given test item matches a single test objective. In the present case, a hypothetical behavior (e.g., raising hand) was analogous to an item, with codes corresponding with test objectives. The following formula was used to index the extent to which the behavior matches a single domain and code:

$$I_{ik} = (N / 2N - 2)(u_k - u)$$

Where N is the number of codes, u_k is the judges' mean rating of behavior *i* on the *k*th code, and *u* is the judges' mean rating on all domains or codes. The highest value of behavior-domain or behavior-code congruence is 1.00, obtained only when an item is matched to a single domain or code by all raters. If the behavior is matched to more than one domain or code, its index will be less than 1.00 and indicates the degree of exclusiveness for each domain or code. Low congruence for a particular behavior suggests that the codes listed by experts for a particular behavior should be refined to ensure that codes are, in fact, mutually exclusive.

Therefore, experts were asked to consider 10 high-frequency behaviors and rate the extent to which they match on each behavioral domain and code. Ratings ranged from -1 (i.e., behavior clearly does not match code and domain) to 1 (i.e., the behavior matches the code and domain). The description of the high-frequency behaviors were drawn from pilot analyses of previous observational studies. Results of this activity indicated that many of the codes were, in fact, mutually exclusive. Of the 10 behaviors presented, seven had a congruence rating of more than .85 suggesting a clear relation with only a single code. One behavior that was supposed to be matched to the *request for attention* code had a rating of .76 indicating a moderate level of congruence and potential for non-mutual exclusivity. In addition, behaviors meant to represent *mild disruptive behavior* (.42) and *behavioral command* (.57) had low congruence ratings suggesting problems with the operational definitions. The low congruence

score associated with *mild disruptive behavior* was determined not to be a concern since all responses from experts for the described behavior were labeled either non-engaged or mild disruptive behavior. Within the context of the coding system, both non-engaged and mild disruptive behaviors could be recorded into the same behavior and used in subsequent analyses. Low congruence scores associated with *behavioral command* led to a revision of the coding scheme to better differentiate command and academic prompt, which was the code most referenced by experts as describing the behavior. This was accomplished by stating that an academic prompt was a direct request to the student (or group of students) to provide a verbal or written response while all other physical responses would be coded as a command.

The third activity for experts to complete was a seven-question Likert exercise designed to assess the adequacy of each specific domain and the ability of the overall coding scheme to capture the range of classroom behaviors emitted by teachers and students. Each item was rated on a five point scale from 1 to 5 (1 = *not at all*; 2 = *somewhat*; 3 = *fairly well*; 4 = *well*; and 5 = *very well*). Responses from the experts indicated that the codes were successful in assessing overall teacher and student functioning as well as subsections of academic and conduct-based behaviors. Table 3 provides the means and standard deviations for the responses across the eight experts sampled.

Table 3.

Expert Ratings on Likert Scale Questions for Content Validity Evaluation

Question	*M*	*SD*
How well do the codes represent teacher behavior during academic periods?	4.00	.58
Do the teacher academic-behaviors represent this construct adequately?	3.86	.38
Do the teacher conduct-based behaviors represent this construct adequately?	4.00	.82
How well do the student codes represent behavior during academic periods?	3.86	.69
Do the student academic behaviors represent this construct adequately?	3.86	.69
Do the behaviors for student conduct-based behaviors represent this construct adequately?[a]	4.20	.45

[a]Only 6 of 8 responses were received to this question due to a misprint in the content validity protocol.

Data collection procedures. The data system was designed and intended to provide an exhaustive representation of all the interactions possible within whole-group instructional contexts between teacher and student. The coding scheme allowed observers to concurrently record specific codes associated with teacher and student behavior. Data was collected using handheld personal digital assistants (PDA) with the Multiple Option Observational System for Experimental Studies (MOOSES; Tapp, Wehby, & Ellis, 1995; Tapp & Wehby, 2000) program uploaded onto each device. The system allowed for the collection of continuous data on the sequence of interactions between the target student and teacher. Specifically, each coded behavior was time stamped onto a digital file and maintained by the computer. As a result, each teacher and student behavior that occurred during the observation session was recorded in a temporal sequence.

Observer Training

A three step process was used to train data collectors to observe and record the identified behaviors. First, data collectors reviewed the coding manual and memorized codes and operational definitions. Questions regarding the differences between codes were answered and minor adjustments to the coding manual were made to increase clarity and understanding across data collectors. Second, data collectors viewed videotaped segments of classroom interactions and practiced inputting data into the PDAs using the MOOSES program. Data collectors had to achieve a minimum of a .80 interrater agreement (i.e., agreement / agreement + disagreement) with a master coding file for each code on three separate occasions on three different videos to move to the third and final stage of training. These master coding files were developed by a senior member of the observational team and were checked independently by

two research assistants with considerable experience collecting observational data in classroom settings. Discussions between the lead author, master coder, and experienced data collectors were used to resolve discrepancies in classifying behaviors on the video clips. Third, qualifying data collectors were then assigned to non-participating classrooms to observe students in naturalistic classroom settings. This allowed observers to adapt to collecting data in classrooms as opposed to videotapes. All observers had an opportunity to collect practice data with different members of the observational team. In order to qualify for data collection, observers had to achieve a minimum of .80 interobserver agreement (i.e., agreement / agreement + disagreement) on all codes with three different observers. Such an arrangement ensured reliable data across several pairs of observers. If an observer was not able to meet any of these conditions, they were moved to another part of the project (e.g., academic assessments). The average interobserver agreement across all codes during these practice sessions was .86 with standard deviation across all codes of .23. Observers demonstrated improvement from the first week of practice sessions (M = .76, SD = .13) to the last (M = .86, SD = .06)

For official data collection each teacher-student dyad was observed four times for 30-minutes per session. A total of 200 data collection sessions were conducted for a total of 6,000 minutes of observation time. All four observation sessions for each teacher-student dyad were pooled for subsequent analyses to create 50 distinct strings of data. Data was collected during whole-class instructional periods with language arts activities being targeted. Teachers were requested to provide the research team with a specific time for language arts prior to the beginning of data collection. Every effort was made by research assistants to collect all four sessions per dyad during language arts periods. In many circumstances, a fifth or sixth session

was added to capture teacher and student behavior during language arts. In such cases, only the four language arts sessions were used for analyses. However, due to scheduling conflicts not all observations were able to be conducted during language arts. Specifically, 10 of the 200 observations (5.00%) were conducted during a period other than language arts. None of the teacher-student dyads had multiple sessions that were non language arts periods. These sessions were not excluded from data analysis because behavioral patterns were similar to those sessions collected during language arts. In addition, each of the non language arts observations was conducted during a language intensive whole-group instructional period such as science or social studies.

Prior to data collection, teachers were asked to complete a *The Class Characteristics and Language Arts Survey* about their language arts periods. This portion of the survey was designed to assess the curricular approaches to language arts used in the classrooms. The variability between classrooms on language arts curriculum was limited with more than two-thirds (83.33%) of teachers reporting the use of MacMillan / McGraw-Hill's Treasures Reading Program (2008). According to the publisher, Treasures is a comprehensive reading program that is based on research evidence. The program attempts to integrate activities related to phonics, vocabulary, reading fluency, comprehension, and writing. In addition, lessons rely on differentiated instruction techniques to assist students with varying abilities learn content. A range of whole-group activities are used within Treasures to promote learning. A typical lesson lasts approximately 45-minutes. Teachers that did not report using a commercial program for language arts instruction tended to rely on activities learned during their teacher preparation courses to guide curricular choices. Approximately one-third (38.00%) of the teachers reported that their school used ability grouping during language arts

instruction. Ability grouping is an instructional approach in which students of similar abilities are placed into the same classroom for one or more academic periods per day. In theory, ability grouping might change the learning environment by allowing the teacher to instruct more effectively since the disparity in student ability levels is reduced. The present study could not address this issue in statistical analyses due to a limited sample size.

Assessing Reliability

A focus of the present study is to validate the use of a coding scheme to measure teacher and student responsiveness. As with any research tool, it is important to demonstrate that data can be collected consistently across individual coders and that each observer can collect data consistently across different conditions (e.g., classrooms, teachers, and students). Therefore, interobserver agreement was estimated using conventional methods to demonstrate that data was collected consistently across observers and settings.

Interobserver agreement refers to the extent to which two or more observers agree on the occurrences and nonoccurrences of a behavior (Suen & Ary, 1989). Indices of interobserver agreement are used to provide evidence that the data are consistent across a number of independent observers. A total of 41 (20.05%) of the observation sessions were used to collect interoberver agreement data. These sessions differed from typical data collection periods by using two, rather than one, observers to collect concurrent data on the teacher-student dyad of interest. Although observers collected data concurrently in the classroom, they were trained to keep data independent from each other. Data collectors were separated to the point that they could not observe the codes entered by the other collector, but they were close enough that they could see and hear the same classroom interactions.

The occurrence agreement index (Bakeman & Gottman, 1997) was used to measure the stability of data between observers. The occurrence agreement index was computed for each individual code and provides information about the percentage of time two observers agree that the behavior of interest occurred or did not occur. The formula used to compute occurrence agreement was:

$$\frac{\text{occurrence agreements}}{\text{occurrence agreements + disagreements}} \times 100\%$$

where "occurrence agreements" are the number of times both observers agree that the behavior occurred and "disagreements" is the number of times the two observers disagree on the occurrence and nonoccurrence of behavior. For the present study, agreement was defined as two independent observers scoring the same code within a 3-second window. The MOOSES program (Tapp et al., 1995) provides a conservative estimate of agreement whereby disagreements between observers are double counted. That is, if one observer recorded a behavior while the other did not the number of disagreements recorded by MOOSES was two. Acceptable levels of occurrence agreement are above 80.00% between observers. The results for interobserver agreement for each code are provided in Table 4. The interobserver agreement for each code ranged from 83% to 96% with an overall agreement of 88%. Overall interobserver agreement was 89% for teacher codes and 84% for student codes.

Table 4

Interobserver Reliability Estimates for Observational Codes

Code Level	Observational Code	Agree	Disagree	Occurrence Agreement
Teacher Behaviors	Instructional Talk	1641	201	.90
	Other Talk	1645	171	.91
	Academic Prompt	1567	196	.88
	Academic Feedback	116	22	.84
	Acknowledgement	50	14	.83
	Command	556	89	.88
Teacher Behaviors	Reprimand	214	27	.89
(cont.)	Praise	147	12	.96
Total		5936	732	.89
Student Behaviors	Engaged	249	40	.91
	Not Engaged	199	60	.87
	Request for Attention	279	37	.87
	Disruptive Behavior	400	83	.90
	Compliant Response	688	123	.85
Total		1815	343	.84
Grand Total		7751	1075	.88

Data Analysis

Measures of sequential association. The operational definitions developed to guide the coding of teacher and student behavior emphasized the sequential nature of teacher-student interactions. Specifically, responsiveness was defined as a reciprocated behavior that was focused on similar content and delivered in a timely fashion. Timeliness was further defined as a behavior emitted by the dyadic partner within a 3-second window. Since data was recorded using continuous timed-event recording methods, the use of time-window sequential analysis was possible for deriving measures of sequential association (Bakeman & Gottman, 1997) Although a full discussion of sequential analytic techniques is beyond the scope of this paper (for more information on sequential analysis the reader is referred to (Bakeman & Gottman), a brief introduction to key terms and concepts will be used to facilitate understanding.

Sequential analysis refers to a set of observational and statistical techniques used to index the magnitude of temporal association between two or more observed events (Gottman & Roy, 1990). Time-window sequential analysis is a subtype of other similar approaches, the most popular of which is lag-event sequential analysis (Bakeman & Gottman, 1997). The main feature that differentiates time-window sequential analysis from other types of sequential analysis is the ability to specify an *a priori* period of time in which the behaviors of interest are expected to occur. For example, a researcher that uses a time-window of 3-seconds expects the consequent behavior to occur within the next three seconds after the occurrence of an antecedent behavior. This approach increases the flexibility that researchers have to measure complex interactions by requiring a specific temporal window to be articulated (Yoder & Tapp, 2004). The results of a sequential analysis provide the researcher with an index of sequential association between the behaviors of interest. Although several methods to

quantifying sequential data have been proposed (Bakeman & Gottman), strength of association (i.e., effect size) measures have proven to be the strongest. The advantage of such measures is their ability to control for random occurrences of the behavioral sequence (Yoder & Feurer, 2000). Two indices of sequential association are often recommended in the sequential analytic literature: the *odds ratio* and its transformation Yule's Q.

The *odds ratio* is the frequency of the antecedent behavior occurring before the consequent behavior over the frequency of the antecedent behavior occurring before all other coded behaviors. A limitation of the *odds ratio* is that it can be difficult to interpret for researchers not accustomed to working with this particular statistic. This is because an *odds ratio* of 1.0 indicates no sequential association between the antecedent and consequent events. No sequential association among behaviors means that the target behavior has an equal chance of occurring after the consequent behavior as a non-consequent behavior. As a result, the occurrence of the given behavior does not increase the probability that the target behavior will occur. In addition, the *odds ratio* is scaled from 0 to infinity. Such asymmetry around 1.0 causes additional interpretation difficulties.

In light of the difficulty that many researchers and non-researchers find with interpreting the *odds ratio*, investigators have suggested a transformation to Yule's Q (Wampold, 1992). The advantages of Yule's Q are identical to that of the *odds ratio*. Namely, it is not influenced by chance occurrences of the behaviors being investigated. By transforming the *odds ratio* to Yule's Q, however, the information contained within the *odds ratio* is more intuitively understood. This is because Yule's Q is scaled from -1.0 to 1.0 with zero indicating no sequential association among behaviors. The strength of association is indicated by its proximity to the absolute value of 1.0. Stronger associations amongst

behavioral events are represented by values closer to |1.0|. A negative Yule's Q means that the consequent behavior follows the antecedent behavior less often than expected by chance processes alone. A positive Yule's Q means that the consequent behavior follows the antecedent behavior more often than expected by chance processes alone. Due to the descriptive advantages associated with the interpretation of sequential data, the use of Yule's Q has been recommended within the sequential analytic literature (Bakeman & Gottman, 1997; Yoder & Tapp, 2004). Therefore, Yule's Q was derived as the measure of teacher and student responsiveness and subsequently used for group comparisons.

Group comparisons. Students at high and low risk for externalizing behavior problems were compared on summary and sequential level measures of teacher and student behavior. First, the groups were compared on the frequencies and durations (i.e., summary level) associated with each code in the coding scheme. These analyses were used to determine if there were any differences between the high and low risk groups on the base rates of teacher behaviors directed toward students and student behaviors. Such information assisted in determining the behavioral context within the classroom and established a framework in which to interpret sequential level data. The second set of analyses compared students at high and low risk on (a) teacher responsiveness to student academic behaviors, (b) teacher responsiveness to student conduct-based behaviors, (c) student responsiveness to teacher academic behaviors, and (d) student responsiveness to teacher conduct-based behaviors. All comparisons of teacher and student responsiveness were based on the Yule's Q computed for each individual dyad.

Both sets of analyses utilized an analysis of variance (ANOVA) framework to compare the frequency and duration of observed behaviors and teacher-student responsiveness. Teacher

behaviors and responsiveness were compared using repeated measures ANOVA in which the two students (i.e., high and low risk) were used as separate conditions. This approach allowed for inherent differences across teachers to be taken into account and for differences in responsiveness toward students at high and low risk to be filtered. Student behaviors and responsiveness were compared using a one-way ANOVA in which risk status was the independent variable. Since there were no statistical differences on academic measures, there was no need to control for reading ability. In addition, an effect size measure (Cohen's *D*) was computed to facilitate understanding of group differences. All effect sizes presented use students at high risk as the referent group

CHAPTER III

RESULTS

Summary Level Analyses

Analysis of the observational data was conducted following initial comparisons of the high and low risk groups on measures of academic and behavioral functioning. The first step was to derive the average frequency and duration of each code for both the high and low risk groups. In addition, the means for each group were compared to identify differences in the amount of teacher and student behavior observed across all four sessions. Results of these analyses are presented in Table 5. In terms of teacher behaviors directed toward students, there were no statistical differences in the durations of instructional talk or other talk as well as the frequencies of observed academic prompts, academic feedbacks, commands, or praise. Teachers were shown to reprimand students at high risk more often than their low risk classmates. In terms of student behaviors, students in the high risk group tended to be academically engaged for less time and displayed more than twice the number of disruptive behaviors than the low risk group. There were no differences between the groups on the frequency of compliant responses or requests for attention.

Table 5

Means, Standard Deviations, and ANOVAs between groups for Teacher and Student Codes

Code Level	Observational Code	High Risk		Low Risk			
		M	SD	M	SD	F	*P*
Teacher Behaviors	Instructional Talk	3786.80	733.04	3544.80	770.01	1.33	.26
	Other Talk	3413.16	732.98	3627.00	935.193	2.24	.15
	Academic Prompt	157.68	63.73	160.44	57.78	.06	.81
	Academic Feedback	21.56	17.61	22.12	12.20	.02	.88
	Command	57.32	27.12	50.84	23.11	3.12	.09
	Reprimand	21.32	14.23	16.36	13.70	13.22	< .01*
	Praise	11.76	11.27	10.08	7.80	1.64	.21
Student Behaviors	Engaged	6268.16	769.12	6805.84	493.60	8.65	< .01*
	Not Engaged	737.88	714.87	289.20	384.838	7.64	< .01*
Student Behaviors	Request for Attention	26.32	22.16	29.88	21.81	.33	.57
	Disruptive Behavior	75.96	51.23	39.52	34.43	8.71	< .01*
	Compliant Response	90.20	47.27	100.76	39.87	.73	.40

Sequential Level Analyses

Teacher Responsiveness. Results comparing teacher responsiveness toward student behavior are presented in Table 6.

Table 6

Means, Standard Deviations, and ANOVAs for Teacher Responsiveness

Student Given	Teacher Response	High Risk		Low Risk		F	*p*	*d*
		M	SD	M	SD			
Compliance	Praise	0.00	.86	-.14	.86	.48	.49	.16
Compliance	Academic Statement[a]	.32	.28	.44	.20	4.82	.04	-.49*
Disruptive	Reprimand	.38	.37	.52	.32	4.07	.05	-.40*
Disruptive	Academic Statement[a]	-.39	.32	-.74	.34	11.16	< .01	1.06*

[a]Academic Statement was defined as providing the student with an academic prompt, academic feedback or the onset of instructional talk following a disruptive behavior.

Recall that each responsiveness measure is presented as a Yule's Q. In the present context, the Yule's Q represents the strength of association between the occurrence of response and given behaviors. Positive values indicate an increased likelihood that the two behaviors co-occur and negative values indicate an inhibiting effect of the given behavior on the target behavior. Stronger relationships are indicated by those that are further away from zero. Also recall that the time-window used to identify teacher and student behaviors was 3-seconds. An appropriate interpretation of these associations would, therefore, include both pieces of information. For

example, there was no association between the provision of teacher praise within 3-seconds following student compliance for high risk students (M = 0.00; SD = .86). The zero-valued Yule's Q indicates that, on average, teacher praise was not contingent on student compliance. In contrast, a small, negative association was observed for teacher praise being emitted within 3-seconds of a compliant response for low risk students (M = -.14; SD = .86). This negative Yule's Q indicates that a compliant response from low risk students actually inhibited the occurrence of teacher praise although this association was generally weak. These findings suggest that teachers were not likely to administer a praise statement within 3-seconds of a compliant response for either high or low risk students. In addition, there were no difference in the sequential association for the high and low risk group F (1, 49) = .48, p = .49. The sequential association between student compliance and teacher instructional statements was also analyzed and compared across groups. For high risk students, a low-moderate association was observed between student compliance followed by teacher academic statement within 3-seconds (M = .32; SD = .28). A high-moderate association was observed between student compliance followed by teacher instructional statement within 3-seconds for the low risk group (M = .44, SD = .20). Mean comparisons between high and low risk groups revealed a statistical difference of teacher academic responsiveness following student compliance F (1, 49) = 4.82, p = .04.

Responsiveness analyses were also conducted for teacher behavior following student disruptive behaviors. Specifically, teachers displayed moderate associations for the provision of a reprimand within 3-seconds of a disruptive behavior for both the high (M = .38; SD = .37) and low risk groups (M = .52; SD = .32). A significant difference was found between the high and low risk groups for the likelihood of teachers to reprimand a disruptive behavior F

(1, 49) = 4.07, p = .05. Significant group differences were also observed between the likelihood of teacher responding with an academic statement to student disruptiveness F (1, 49) = 11.16, p = < .01. In this case, disruptiveness was shown to have an inhibiting impact on teacher academic statements for both groups. That is, teacher instruction was less likely to occur following a disruptive behavior with the strength of association being stronger for the low risk group (M = -.74, SD = .34) than for the high risk group (M = -.39, SD = .32).

Student Responsiveness. Results of the sequential level analyses for student responsiveness to teacher behavior are presented in Table 7. In terms of student responses to teacher academic statements, students at low risk (M = .75, SD = .09) were more likely to emit an academic response within 3-seconds of a teacher academic behavior than high risk students (M = .67; SD = .17). Group comparisons for this association were statistically significant F (1, 49) = 4.38, p = .04. No association was observed between the academic statements of teachers and the emitting of student disruptive behaviors within 3-seconds for the high risk group (M = 0.00; SD = .33) whereas there was a weak inhibiting effect for low risk students (M= -.22, SD = .43). This was shown to be statistically different according to group comparisons F (1, 49) = 4.12, p = .05.

Table 7

Means, Standard Deviations, and ANOVAs for Student Responsiveness

Teacher Given	Student Response	High Risk M	High Risk SD	Low Risk M	Low Risk SD	F	*p*	*d*
Academic Statement[a]	Compliance	.67	.17	.75	.09	4.38	.04	-.59*
Academic Statement[a]	Disruptive	0.00	.33	-.22	.43	4.12	.05	.57*
Command	Compliance	.76	.16	.81	.08	2.19	.14	-.40
Command	Disruptive	-.18	.39	-.36	.50	1.78	.19	.40
Reprimand	Compliance	.03	.63	.31	.65	3.41	.07	-.44
Reprimand	Disruptive	-.54	.54	-.75	.54	1.87	.18	.39

[a]Academic Statement was defined as providing the student with an academic prompt, academic feedback or the onset of instructional talk following a disruptive behavior.

Although there were no significant group differences found for the association of teacher command with either student compliant $F = (1, 49)\ 2.19$, $p = .14$ or disruptive behavior $F = (1, 49)\ 1.78$, $p = .19$, students at low risk were shown to have a stronger association with compliant responses and have their disruptiveness inhibited to a greater degree. Students at low risk tended to emit a compliant response within 3-seconds of a teacher reprimand ($M = .31$, $SD = .65$) more often than high risk students ($M = .03$, $SD = .63$). Group comparisons were not significant though the association of student compliance following teacher reprimand was approaching significance $F\ (1, 49) = 3.41$, $p = .07$. Finally, students at low risk ($M = .22$, $SD = .57$) tended to have their disruptive behaviors inhibited to a greater

degree than their high risk classmates (M = .61, SD = 1.09). However, these differences were not statistically significant F (1, 49) = 1.87, p = .18.

CHAPTER IV

DISUCSSION

Research on school based behavioral problems has traditionally focused on relationships that act in a unidirectional manner (Sutherland & Oswald, 2005). That is, a majority of research has considered only the effect of teacher behavior on student behavior (e.g., Jolivette, Wehby, Canale, & Massey, 2001; Madsen, Becker, & Thomas, 1968; Sutherland, Alder, & Gunter, 2003) or the impact that student behavior has on teacher behavior (Alber & Heward, 2000). However, a more complete understanding of classroom functioning should consider the reciprocity between teacher and student behavior. Improvements to traditional modeling and measurement techniques of teacher and student behavior can provide researchers with a greater understanding of the complexity of classroom based interactions. The present study was an attempt to extend the present literature on teacher-student interactions by comparing the interactional sequences of teachers and students at high and low risk of developing externalizing behavior problems.

The findings of this study revealed few differences in the summary level data for both teachers and students. Although the high and low risk groups were shown to be similar on many of the summary level variables, there are some important descriptive points to be made. For example, teachers spent about half of the observation sessions engaged in academic instruction toward the whole class or individual target students. Within this seemingly low duration of overall instruction, teachers provided drastically fewer academic prompts and praise statements than the literature recommends is necessary to maintain the engagement

levels of students with EBD (Council for Exceptional Children, 1987). In addition, there were low rates of academic feedback observed to both high and low risk students. The only summary level teacher behavior in which a difference between high and low risk students was observed was the total number of reprimands where high risk students were shown to be reprimanded nearly twice as much as low risk students. These findings on the summary level behaviors of teachers are consistent with results of past studies (e.g., Shores et al., 1993; Wehby et al., 1995).

In terms of the summary level data for student behavior, there were no differences in the observed rates of student request for attention or compliance. Although the base rates of these behaviors were not statistically significant across groups, it is the proximity of such responses to teacher behavior that dictated the degree of observed responsiveness. Differences between the high and low risk groups were observed for time spent academically engaged and the number of disruptive behaviors emitted. Specifically, both groups demonstrated high rates of engagement with high risk students having an average engagement time of 6,268.16 minutes (87.06%) and low risk students having an average engagement time of 6,805.94 minutes (95.04%). In terms of disruptive behaviors, students at high risk were shown to emit considerably more than low risk students. In essence, this finding demonstrates that high risk students were probably more challenging for teachers to instruct and manage than low risk students.

According to the summary level data, there was little evidence that positive social consequences by the teacher (i.e., praise) were given for appropriate student behaviors (i.e., compliance). On the surface, the woefully low frequency of observed praise suggests that teachers rarely attended to the positive behaviors of students. However, such findings might

have been misleading if praise statements were delivered in such a way as to promote desirable student behavior. The sequential level data suggested, however, that teacher praise was not contingent on positive student behaviors for high risk students and actually was inhibited by positive behaviors of low risk students. Therefore, it seems as though teachers in this sample provided praise at seemingly random times. According to operant theory (Catania, 2007), this lack of responsiveness places compliant behaviors on extinction thereby reducing the likelihood of compliance over time. Student compliance is, therefore, undermined by the teacher's incongruent (or non-existent) responsiveness. The failure to re-establish the contingency between prosocial classroom behaviors and positive consequences probably has a greater impact on students predisposed to disruptive behaviors such as those at high risk for EBD.

A more common teacher response to student compliance was the provision of an academic statement. The sequential associations between an academic statement made by the teacher and student compliance were moderate for both groups. However, teachers were more likely to direct instruction toward low risk students than to high risk students following a compliant response. The increased likelihood that teachers provided an academic statement to low risk students can be understood in terms of coercion theory. Within this framework, coercive interactions in the classroom are predicated on a cycle of negative reinforcement whereby student responses shape the behavior of teachers and vice versa. It was shown that both students at high and low risk were compliant to the academic statements of teachers. However, low risk students were more likely to be compliant. In addition, the academic statements of teachers tended to have an inhibiting effect on the occurrence of student disruptiveness for low risk students while there was no association demonstrated for high risk

students. The fact that teachers were more likely to withdraw instruction from high risk students is likely related to the observed contingencies following teacher instruction. Specifically, low risk students were (a) more likely to show compliance with academic statements and (b) more likely to have their disruptive behaviors inhibited following teacher academic statements. Though the presence of coercion may not be evident through the examination of high-risk data alone, when compared to the responsiveness of low risk students, patterns of coercion were detected.

Similar to the analyses involving prosocial behaviors, coercive interactions (Gunter et al.,1993, 1994;Patterson, 1982) between teachers and students were observed for both groups surrounding antisocial behaviors as well. The mean association between student disruptive behaviors and teacher reprimand was large for students at high and low risk. Although students from the high risk group were shown to emit more than twice as many disruptive behaviors and received nearly twice as many reprimands during observations, the sequential level analyses revealed that teachers were more likely to reprimand low-risk students within 3-seconds following a disruptive behavior. Teachers were clearly responsive to the disruptiveness of students from both groups. However, the increased rate of responding toward low risk students in this case suggests a possible desensitization to the disruptiveness of high risk students. In fact, the association between teacher reprimand and student compliance was trending toward no association for high risk students and a moderate association for low risk students. Taken together, it seems as though teachers might have been less responsive to the disruptiveness of high risk students because they were less likely to comply. Within the model of coercion (Patterson 1976, 1982; Gunter et al., 1993, 1994), the removal of teacher consequences for student disruptive behavior would reinforce the teacher by avoiding an

increased chance of non-compliance. Although there were no perceptible group differences in the mean sequential associations between (a) teacher command and student compliance, (b) teacher command and student disruptive, and (c) teacher reprimand and student disruptive, the trend of each of these measures lends further support to the possibility of coercion between teachers and students at high risk for EBD. Specifically, low-risk students were more likely to comply with teacher commands, less likely to disrupt following teacher commands, and less likely to be disruptive following teacher reprimands.

Summary of Findings

In general, the sequential level results reported in this study provide a more thorough illustration of the dynamic micro-social processes within teacher-student interactions than summary level analyses alone. The findings of the present study seem to suggest that students at high risk for EBD engage in more coercive interactions with their teachers than low risk students. Elements of coercion were found for both antisocial and prosocial behaviors through the investigation of teacher and student responsiveness. Specifically, teachers tended to respond with fewer academic statements to compliant high risk students as compared to compliant low risk students. The underlying reasons for this difference might be partially attributable to observed levels of student responsiveness to teacher academic statements. That is, high risk students were less likely to be compliant and less likely to have their disruptiveness inhibited. A similar pattern of responding was found for antisocial behaviors. Teachers were more likely to reprimand low risk students within 3-seconds, which on the surface might be surprising. However, high risk students tended to be less compliant and more disruptive following teacher commands and reprimands. Therefore, the interactions between

teachers and high risk students tend to follow a pattern of greater coerciveness as compared to those of low risk students.

Limitations

The results from the present study should be considered with the following limitations in mind. First, the primary goal in the development of the coding scheme was to account for the range of teacher and student behaviors emitted during whole-group instruction. Although the coding scheme was subjected to a content evaluation by experts in the field of observational methods with results indicating a sound measurement tool, some important teacher and student behaviors might have been overlooked. In addition, some observational research has indicated that peer behavior can impact the onset and maintenance of problem behavior in classroom settings (Shore et al., 1993; Taylor & Romancyk, 1994). Since the focus of the present study was to develop a measurement tool of teacher-student interactions, peer behaviors were purposefully not included in the present coding scheme. The sampling of high and low risks students from the same classroom was an attempt to control for classmate behavior and other contextual factors that might impact teacher and student behavior.

A second limitation of the present study was the inability to measure the function of behavior for high and low risk students. The focus of this study was on differences in the patterns responding by teachers and students at high and low risk for EBD. Although students might demonstrate similar behavioral topographies in response to certain teacher behaviors (both appropriate and inappropriate), the purpose for emitting these response patterns might vary across individuals. For example, one student might be disruptive to access attention from the teacher while another might want to escape academic instruction. Teacher and student

responding might have been impacted by the function of student behavior in addition to the topography. Additional research is required to determine the extent to which the function of student behavior influences the level of coercion and response patterns of teachers.

A third limitation of this study was the limited sample size. Prior to the study, a power analysis was conducted to determine the minimum sample required to detect group differences. Studies that compared various operationalizations of teacher-student interactions across students with different behavioral profiles (e.g., students at high risk for EBD vs. students at low risk; students with considered to be high aggressors vs. low aggressors; students with high and low levels of academic engagement) were sampled. Results of the power analysis revealed that a sample size of approximately 25 dyads was required to have an 80% chance of finding group differences at the .05 probability level. Since 25 teacher-student dyads were eligible for participation in the present study, this represents a sample size that was toward the bottom of the threshold to find effects. This should be taken into consideration particularly for sequential level variables that were approaching significance. Furthermore, the present sample size was not sufficient to determine if patterns of coercion increased across grade. According to referral data, students at-risk for EBD tend to be identified at higher rates toward the middle of elementary school (Kauffman, 2005). If patterns of coercion between teachers and students at-risk for EBD become more evident from 1st to 3rd grades, such data might have implications for identification and assessment.

A fourth limitation of this study is its descriptive nature. Descriptive studies are used to describe naturally occurring phenomena (Gall, Gall, & Borg, 2007). A drawback of descriptive studies is a reliance on correlations between different variables to draw conclusions rather than demonstrating an effect of one variable on another through

experimentation. Results of correlational studies must, therefore, be interpreted carefully. Conclusions between variables cannot be understood as causal but as a description of the relationship among variables for the given sample. Though correlational studies do not provide adequate grounds for determining causality, the information derived from such research can be valuable for developing effective, empirically based interventions (Gall et al.).

Implications for Research and Practice

The findings of the present study have implications for both research and practice. In terms of research, future investigations comparing the responsiveness of teachers and students at risk for EBD should focus on one or more of the following four areas. First, future research on teacher-student interactions should attempt to integrate current theoretical perspectives to develop a comprehensive model of micro-social processes in the classroom. Areas of interest might include, but are not limited to: the impact of various contextual factors on the classroom behavior of teachers and students; the variation in behavioral topographies associated with different behavioral functions; and the deterioration of teacher-student interactions over time. Time might be conceptualized as either within a single school year or across grade level given the evidence that students with EBD tend to be identified at higher rates in the 3rd and 4th grades. A second topic for subsequent research is to address methodological improvements to the study of teacher-student interactions. Further development and validation of comprehensive coding schemes to collect reliable data on the classroom behavior of teachers and students would provide better measurement of teacher and student behaviors. In addition, studies concerned with the amount of observational data required to make reliable inferences would benefit the field. The third area for future research is to demonstrate a functional

relation between teacher responsiveness and student behavior. Demonstrating a functional relation between teacher behavior and student behavior can lay the groundwork for future intervention research similar to those currently found in the parent-child literature (e.g., Heriot, Evans, & Foster, 2001). Coupled with better theoretical understanding of the contextual factors impacting classroom behavior and improvements to the current methodological limitations, such interventions might result in improved school experiences for children with and at-risk for EBD.

The current study also has implications for practice. Results indicated that teachers tended to respond differently to high and low risk students despite the presence of similar behavioral topographies. In addition, it was found that teachers often fail to praise appropriate behavior for either the high or low risk groups. Unfortunately, the response patterns identified here might impact the manner in which students, particularly those at high risk for EBD, behave in classroom settings. Although additional research is required to develop a comprehensive intervention model for reducing the level of coercion in teacher-student interactions, there are a number of low-cost techniques to assist teachers monitor their instructional and management behaviors toward students. Examples of relevant interventions include self-monitoring (e.g., Kalis, Vannest, & Parker, 2007) and teacher-feedback interventions (e.g., Sutherland & Wehby, 2001; Sutherland, Wehby, & Copeland, 2000). Current practitioners should strongly consider employing such interventions in conjunction with recommended benchmarks (e.g., Council for Exceptional Children, 1987) to modify their classroom behavior. In the meantime, research will continue to utilize theory and data from both descriptive and experimental studies to develop contextually bound interventions that are flexible, effective, and maximize teacher time and resources.

REFERENCES

Achenbach, T. M & Rescorla, L. A. (2001). *Manual for the Achenbach System of Empirically Based Assessment*. Burlington: University of Vermont, Department of Psychiatry.

Alber, S. R., & Heward, W. L. (2000). Teaching students to recruit positive attention: A review and recommendations. *Journal of Behavioral Education, 10,* 177-204.

Anderson, J. A., Kutash, K., & Duchnowski, A. J. (2001). A comparison of the academic progress of students with EBD and students with LD. *Journal of Emotional and Behavioral Disorders, 9,* 106-115.

Bakeman, R. & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). New York: Cambridge University Press.

Baker, J. A. (1999). Teacher student interactions in urban at-risk classrooms: Differential behavior, relationship quality, and student satisfaction with school. *The Elementary School Journal, 100,* 57-70.

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

Carr, E. G., Taylor, J. C., & Robinson, S. (1991). The effects of severe problem behavior in children on the teacher behavior of adults. *Journal of Applied Behavioral Analysis, 24,* 523-535

Catania, A. C. (2007). *Learning* (4th ed.). Cornwall-on-Hudson, NY: Sloan Publishing.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Council for Exceptional Children. (1987). Academy for effective instruction: Working with mildly handicapped students. Reston, VA: Author.

Criss, M. M., Shaw, D. S., & Ingoldsby, E. M. (2003). Mother-son positive synchrony in middle childhood: Relation to antisocial behavior. *Social Development, 12,* 379-400.

Deater-Deckard, K., Atzaba-Poria, N., & Pike, A. (2004). Mother – and father – child mutuality in Anglo and Indian British families: A link with lower externalizing problems. *Journal of Abnormal Child Psychology, 32,* 609-620.

Deater-Deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Pscyhopathology, 10,* 469-493.

Drummond, T. (1994). *The Student Risk Screening Scale (SRSS).* Grants Pass, OR: Josephine County Mental Health Program.

Drummond, T., Eddy, J. M., & Reid, J. B. (1998a). *Follow-up study #3; risk screening scale: Prediction of negative outcomes by 10th grade from 2nd grade screening.* Unpublished technical report. Eugene, OR: Oregon Social Learning Center.

Drummond, T., Eddy, J. M., & Reid, J. B. (1998b). *Follow-up study #4; risk screening scale: Prediction of negative outcomes in two longitudinal samples.* Unpublished technical report. Eugene, OR: Oregon Social Learning Center.

Drummond, T., Eddy, J. M., Reid, J. B., & Bank, L. (1994, November). *The student risk screening scale: A brief teacher screening instrument for conduct disorder.* Paper presented at the Fourth Annual Prevention Conference, Washington, DC.

Echevarria, J., Fisher, D., Gibson, V., Hasbrouck, J., Shanahan, T. (2008). *Treasures: A reading / language-arts program*. New York: Macmillan / McGraw-Hill.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). New York: Allyn & Bacon.

Gottman, J. M. & Roy, A. K. (1990). *Sequential analysis: A guide for behavioral research.* New York: Cambridge Press.

Gunter, P. L., Denny, R. K., Jack, S. L., Shores, R. E., & Nelson, C. M. (1993). Aversive stimuli in academic interactions between students with serious emotional disturbance and their teachers. *Behavioral Disorders, 18,* 265-274.

Gunter, P.L., Shores, R.E., Jack, S.L., Denny, R.K., & DePaepe, P.A. (1994). A case study of the effects of altering instructional interactions on the disruptive behavior of a child identified with severe behavior disorders. *Education and Treatment of Children, 17,* 435- 444.

Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental Review, 22,* 555-592.

Harrison, J. S., Gunter, P. L., Reed, T. M., & Lee, J. M. (1996). Teacher instructional language and negative reinforcement: A conceptual framework for working with students with emotional and behavioral disorders. *Education and Treatment of Children 19,* 183-196.

Heriot, S. A., Evans, I. M., & Foster, T. M. (2001). An interactional approach to intervention research with children diagnosed with ADHD. *Journal of Child and Family Studies, 10,* 287-299.

Herring, M., & Wahler, R. G. (2003). Children's cooperation at school: The comparative influences of teacher responsiveness and the children's home-based behavior. *Journal of Behavioral Education, 12,* 119-130.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127-155.

Isabella, R. A., & Belsky, J. (1991). Interactional synchrony and the origins of infant-mother attachment: A replication study. *Child Development, 62,* 373-384.

Jolivette, K., Wehby, J.H., Canale, J., & Massey, N.G. (2001). Effects of choice making opportunities on the behavior of students with emotional and behavior disorders. *Behavioral Disorders, 26,* 131-145.

Kalis, T. M., Vannest, K. J., & Parker, R. (2007). Praise counts: Using self-monitoring to increase effective teacher practices. *Preventing School Failure, 51,* 20-27.

Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). What is special about special education for children with emotional or behavioral disorders? *The Journal of Special Education, 37,* 148-156.

Lane, K. L. (1999). Young students at risk for antisocial behavior: The utility of academic and social skills intervention. *The Journal of Emotional and Behavioral Disorders, 7,* 211-223.

Lane, K. L., Little, M. A., Casey, A. M., Lambert, W., Wehby, J., Weisenbach, J. L. et al. (2008). A comparison of systematic screening tools for emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 17,* 178-190.

Lindsey, E. W., Colwell, M. J., Frabutt, J. M., & MacKinnon-Lewis, C. (2006). Family conflict in divorced and non-divorced families: Potential consequences for boys' friendship status and friendship quality. *Journal of Social and Personal Relationships, 23,* 45-63.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., Van Kammen, W. B. (1998). Multiple risk factors for multiproblem boys: Co-occurrence of delinquency, substance use, attention deficit, conduct problems, physical aggression, covert behavior, depressed mood, and shy/withdrawn behavior. In R. Jessor (Ed.). *New Perspectives on Adolescent Risk Behavior* (pp. 90-149). New York: Cambridge University Press.

Madsen, C. H., Becker, W. C., Thomas, D. R. (1968). Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis, 1,* 139-150.

McIntosh, K., Horner, R. H., Chard, D. J., Dickey, C. R., & Braun, D. H. (2008). Reading skills and function of problem behavior in typical school settings. *The Journal of Special Education, 42,* 131-147.

Montague, M. & Rinaldi, C. (2001). Classroom dynamics and children at risk: A followup. *Learning Disability Quarterly, 24,* 75-83.

Moore, S. R. & Simpson, R. L. (1983). Teacher-pupil and peer verbal interactions of learning disabled, behavior-disordered, and nonhandicapped students. *Learning Disability Quarterly, 6,* 273-282.

Nelson, R. J. & Roberts, M. L. (2000). Ongoing reciprocal teacher-student interactions involving disruptive behaviors in general education classrooms. *Journal of Emotional and Behavioral Disorders, 8,* 27-37.

Patterson, G. R. (1976). The aggressive child: Victim and architect of a coercive system. In L. A. Hamerlynck, E. J Mash, & L. C. Handy (Eds.), *Behavior modification and families: Theory and research (Vol. 1).* New York: Brunner / Mazel, 1975

Patterson, G. R. (1982). *Coercive family interactions.* Eugene, OR: Castalia Press.

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disorders. *The Journal of Special Education, 38,* 130-144.

Rocissano, L., & Yatchmink, Y. (1983). Language skill and interactive patterns in prematurely born toddlers. *Child Development, 54,* 1229-1241.

Rovinelli, R. J., & Hambleton, R. K. (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Dutch Journal of Educational Research, 2,* 49-60.

Sabornie, E. J., Cullinan, D., Osborne, S. S., & Brock, L. B. (2005). Intellectual, academic, and behavioral functioning of students with high-incidence disabilities: A cross-categorical meta-analysis. *Exceptional Children, 72,* 47-63.

Severson, H., & Walker, H. M., (2002). Developmental prevention of at-risk outcomes for vulnerable antisocial children and youth. In K. L. Lane, F. M. Gresham, & T. E. O'Shaughnessy (Eds.), Interventions for children with or at risk for emotional and behavioral disorders (pp. 177–194). Boston: Allyn & Bacon.

Shores, R. E., Jack, S. L., Gunter, P. L., Ellis, D. N., DeBriere, T. J., & Wehby, J. H. (1993). Classroom interactions of children with behavior disorders. *Journal of Emotional and Behavioral Disorders, 1,* 27-39.

Skiba, R. J., Poloni-Staudinger, L., Gallini, S., Simmons, A. B., & Feggins-Azziz, R. (2006). Disparate access: Disproportionality of African American students with dsiabilities across education environments. *Exceptional Children, 72,* 411-425.

Skinner, E. A. & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571-581.

Stormont, M. (2002). Externalizing behavior problems in young children: Contributing factors and early intervention. *Psychology in the Schools, 39,* 127-138.

Suen, H. K. & Ary, D. (1989). *Analyzing quantitative behavioral observation data.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Sutherland, K. S., Alder, N., & Gunter, P. L. (2003). The effect of varying rates of opportunities to respond to academic requests of the classroom behavior of students with EBD. *Journal of Emtotional and Behavioral Disorders, 4,* 239-248.

Sutherland, K. S. & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies, 14,* 1-14.

Sutherland, K. M., & Wehby, J. (2001). The effects of self-evaluation on teacher behavior in classrooms for students with emotional and behavioral disorders. *Journal of Special Education, 35,* 161-171.

Sutherland, K. M., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8,* 2-8.

Tapp, J., & Wehby, J. (2000). Observational software for laptop computers and optical bar code time wands. In T. Thompson, D. Felce, & F. Symons (Eds.), Behavioral observation: Technology and applications in developmental disabilities (pp. 71–82). Baltimore: Brookes.

Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, and Computers, 27,* 25-31.

Taylor, J. C. & Romancyzk, R. G. (1994). Generating hypotheses about the function of student problem behavior by observing teacher behavior. *Journal of Applied Behavior Analysis, 27,* 251-265.

Van Acker, R., Grant, S. H., & Henry, D. (1996). Teacher and student behavior as a function of risk for aggression. *Education and Treatment of Children, 19,* 316-334.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in schools: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth.

Walker, H. M., Shinn, M. R., O'Neil, R. E., & Ramsey, E. (1987). Longitudinal assessment and long-term follow-up of antisocial behavior in fourth-grade boys: Rationale, methodology, measures, and results. *Remedial and Special Education, 8*, 7-16.

Wampold, B. E. (1992). The intensive examination of social interaction. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 93–132). Hillsdale, NJ: Erlbaum.

Wehby, J. H., Symons, F. M., Canale, J., & Go, F. (1998). Teaching practices in classrooms for students with emotional and behavioral disorders: Discrepancies between recommendations and observations. *Behavioral Disorders, 24*, 52-57.

Wehby, J. H., Symons, F. M., & Shores, R. E. (1995). A descriptive analysis of aggressive behavior in classrooms for children with emotional and behavioral disorders. *Behavioral Disorders, 20*, 87-105.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Pulishing.

Yoder, P. J. & Feurer, I. D. (2000). Quantifying the magnitude of sequential association between events or behaviors. In F. J. Symons, T. Thompson, D. Felce, *Behavioral observation: Technology and applications in developmental disabilities*, 317-333. Baltimore MD: Paul H Brookes Publishing.

Yoder, P. J. & Tapp, J. Empirical guidance for time-window sequential analysis of single cases. *Journal of Behavioral Education, 13*, 227-246.