Using a multi-level model to examine the fidelity of implementation of school-wide positive behavior intervention support and its relationship to academic achievement in Louisiana

Michelle Farnsworth Botos
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Education Commons

Recommended Citation
Botos, Michelle Farnsworth, "Using a multi-level model to examine the fidelity of implementation of school-wide positive behavior intervention support and its relationship to academic achievement in Louisiana" (2013). LSU Doctoral Dissertations. 1124.
https://digitalcommons.lsu.edu/gradschool_dissertations/1124

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
USING A MULTI-LEVEL MODEL TO EXAMINE
THE FIDELITY OF IMPLEMENTATION OF SCHOOL-WIDE
POSITIVE BEHAVIOR INTERVENTION SUPPORT AND ITS
RELATIONSHIP TO ACADEMIC ACHIEVEMENT IN LOUISIANA

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Educational Theory, Policy, and Practice

by
Michelle Lynn Farnsworth Botos
B.A., Louisiana Tech University, 2001
M.A.T., Southeastern Louisiana University, 2005
May 2013
I dedicate my dissertation to my wonderful family, who have all provided continuous support to me on this journey. I could not have asked for more encouragement, help, love, patience, and support along this road. To my loving parents, Clark and Myra Farnsworth, who instilled in me a love, passion, and desire to be all that I could be. You have been such an encouragement and infinite help. You were always there to provide just the words that I needed to hear and were there to take all of the kids wherever they needed to be so I could write. Most importantly, you have both always motivated, supported, and believed in me to accomplish one of my life-long dreams. To my sister and brothers, Pamela, Chris, and Kevin, you all have been my biggest cheerleaders and shown me the importance of hard work, all while adding a little humor when I needed it most. To my mother and father-in-law who have always stood beside me. Thank you for welcoming me into your family and always giving me the support to see my goals to completion. To my four precious children, Marcus, Mallory, Addisyn, and Brennen, who have sacrificed so much to allow me to complete my education. Thank you so much for understanding when I spent countless hours in class or on schoolwork when I wished I could have been spending time with you all. I hope that I have been an example to show you that anything you can dream of, you can achieve! Thank you each for being the biggest blessings in my life. And most of all to my precious husband Ben who has walked with me step-by-step, encouraged me when I was ready to give up, and given me more support than could ever have been expected. You truly have made this dream become a reality for me. I thank each and every one of you for giving so much of yourselves so that I could fulfill my educational goals. I could not have done this without each and every one of you.
ACKNOWLEDGEMENTS

I would like to thank my committee members who helped me so much with their endless support and time invested. To Dr. R. Kenton Denny, my committee chairman for his expertise, reflecting, reading, and most of all his continued support throughout this process. A special thank you to Dr. Eugene Kennedy for his countless hours of meeting, reading, guiding, statistical expertise, and most of all for empowering me with so much knowledge. Dr. Kennedy, you went over and beyond in all areas to guide me through this process and you always believed in me. I thank you from the bottom of my heart. Thank you to Dr. Earl Cheek, Dr. Paul Mooney, and Dr. Maribel Dietz for agreeing to serve on my committee. I thank them for their contribution and support to make my research better. Also a special thank you to both Dr. Kara Hill and Mrs. Wendy Allen, my colleagues and more importantly friends. Kara and Wendy, you both were always there to guide me through this process with your support and availability. You have made me laugh and encouraged me at every turn.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... iii

LIST OF TABLES ......................................................................................................................... vi

ABSTRACT ..................................................................................................................................... vii

CHAPTER 1: INTRODUCTION ...................................................................................................... 1

  Context of the Problem ............................................................................................................... 6
  Purpose of the Study .................................................................................................................... 6

CHAPTER 2: LITERATURE REVIEW .......................................................................................... 9

  History of PBIS .......................................................................................................................... 11
  Systems of PBIS ........................................................................................................................ 12
  Implementation and Academic Achievement ............................................................................. 15
  Academic Achievement and PBIS ............................................................................................ 16
  State-wide PBIS and Academic Achievement ......................................................................... 21
  PBIS in Louisiana ...................................................................................................................... 24

CHAPTER 3: METHODOLOGY ................................................................................................... 27

  Multilevel Model ....................................................................................................................... 27
  Sample and Population ............................................................................................................. 30
  Measures/Variables .................................................................................................................. 31
  Operational Definitions ............................................................................................................ 32
  Achievement Scores ................................................................................................................ 32
  Student Level Variable ............................................................................................................ 33
  School Level Variables ............................................................................................................ 33
  Analysis Strategy ...................................................................................................................... 34
  The Model ............................................................................................................................... 37

CHAPTER 4: DATA ANALYSIS ................................................................................................ 40

  Data Quality ............................................................................................................................. 41
  Data Comparison ...................................................................................................................... 42
  Research Question #1 ............................................................................................................... 42
  Research Question #2 ............................................................................................................... 46
  The Model ............................................................................................................................... 48
  Research Question #3 ............................................................................................................... 50
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS .............................................................. 60

Conclusions, Implications, and Recommendations ............................................................ 63

REFERENCES .......................................................................................................................... 69

APPENDIX A: LIST OF ACRONYMS/ABBREVIATIONS USED IN CURRENT RESEARCH ....... 76

VITA ........................................................................................................................................ 79
LIST OF TABLES

Table 1. Spring 2008 LEAP Scores for Site Code 032003 on Data Set ........................................... 41
Table 2. Spring 2008 LEAP Scores for Site Code 032003 on School Report Card ..................... 41
Table 3. English Language Arts: Number of Students at Each Achievement Level ................. 42
Table 4. Demographics for Data Set with All 3 Years of Test Scores ........................................... 43
Table 5. Pearson Correlation for High SET Score, 2007 BOQ, and 2008 BOQ ......................... 44
Table 6. Summary of Correlations ................................................................................................. 45
Table 7. Correlations and Descriptive Statistics for Grade 3 to Grade 4 Match .................... 48
Table 8. Correlations and Descriptive Statistics for the Grade 3 to Grade 4 Match Only Including Students with SET School Scores .................................................................................. 49
Table 9. Correlations and Descriptive Statistics for Grade 4 to Grade 5 Match .................... 50
Table 10. Correlations and Descriptive Statistics for the Grade 4 to Grade 5 Match Only Including Students with SET School Scores .................................................................................. 50
Table 11. Unconditional Model for Grade 3 to Grade 4 and Grade 4 to Grade 5 Match .......... 51
Table 12. Reading: Grade 3 to Grade 4 Using High SET Score .................................................... 52
Table 13. Math: Grade 3 to Grade 4 Using High SET Score ......................................................... 53
Table 14 ELA: Grade 3 to Grade 4 Using High SET Score ............................................................ 54
Table 15. Science: Grade 3 to Grade 4 Using High SET Score .................................................... 54
Table 16. Social Studies: Grade 3 to Grade 4 Using High SET Score ........................................ 55
Table 17. Reading: Grade 4 to Grade 5 Using High SET Score .................................................... 56
Table 18. Math: Grade 4 to Grade 5 Using High SET Score ........................................................ 57
Table 19. ELA: Grade 4 to Grade 5 Using High SET Score ........................................................ 57
Table 20. Science: Grade 4 to Grade 5 Using High SET Score .................................................... 58
Table 21. Social Studies: Grade 4 to Grade 5 Using High SET Score ........................................ 58
ABSTRACT

This study is to examine if implementation of the universal level of PBS is related to student achievement on the LEAP and iLEAP examinations administered as part of the Louisiana Educational Assessment Program. A second purpose is to examine whether identified student/school characteristics contribute to any difference in academic performance. To address this question, it is necessary to take into consideration individual as well as school level factors that may act to facilitate or impede student achievement. Multilevel statistical models are ideally suited for research problems of this nature and will be the approach taken for this study.

The present study sought to determine if a high level of implementation of PBIS at the school level had an impact on a student’s academic scores based on high stakes test scores. Across the state of Louisiana, a schools’ implementation level of PBIS did show evidence of change of a student’s high-stakes test score from 2007 to 2008. However, it did not show a change from 2008 to 2009. In both 2 year spans, a previous year’s academic test score and a school’s percent of free and reduced lunch were both predictors in the model. In 2007-2008, a school’s SET score was also a significant predictor in the model in all of the content areas, except for math. The findings from the multilevel model are consistent with other findings for the 3rd grade to 4th grade match. However, for the 4th to 5th grade match the absence of similar findings may reflect the fact that implementation scores were mandated at the state level for these years, which may have caused inflation in scores from each school. Even in the correlations, the 2008 scores were not statistically significant as the 2007 scores were found to be.
CHAPTER 1: INTRODUCTION

School administrators and teachers across the country are increasingly confronted with disruptive behavior by students (Sugai, Sprague, Horner, & Walker, 2000). Lassen, Steele, and Sailor (2006) state, “In addition to the responsibility of teaching academic subjects such as math, reading, science, the arts, and writing, educators must increasingly deal with nonacademic factors that influence the instruction they provide.” Student disruptive behavior, according to Sugai and Horner (1994), consume an ever growing proportion of teacher and school resources.

Many schools rely on some form of exclusion and zero tolerance policies to control the behavior of students who commit serious offenses such as bringing a weapon or drugs to schools (Skiba & Knesting, 2001). Traditional disciplinary methods, which include detentions, in-school suspension, out of school suspension, and expulsion, have been found to be ineffective and may only exacerbate problem behaviors (Lewis & Garrison-Harrell, 1999; Lewis-Palmer, Sugai, & Larson, 1999). Schools often react to minor offenses with exclusionary practices in lieu of a positive, proactive approach to student misbehavior, often with unintended consequences, such as those unwanted behaviors actually increasing (Skiba & Knesting, 2001).

In response to the increase of antisocial and disruptive behavior by students, schools have created “get tough” and “zero tolerance” approaches to address misbehavior; however, these reactive and punitive approaches to challenging student behavior have been criticized as a short-term solution and leave out an important function of schools— teaching (Noam, Warner, & Van Dyken, 2001). Unfortunately, the success reactive and punitive strategies have
not been sufficiently studied, and some researchers even suggest that these procedures can even increase problem behavior (Shores, Gunter, & Jack, 1993). Included in these adverse strategies, suspensions and expulsions have been linked to academic failure, higher dropout rates, and increased criminal activity (Skiba & Peterson, 2000).

In schools, the key to effective prevention of these negative behaviors is to alter environments and the adults’ behaviors, which result in positive outcomes for students (Scott, Alter, Rosenberg, & Borgmeier, 2010). While maintaining safety at school is paramount, students who are disadvantaged, minority, or have disabilities are the recipients of suspensions and expulsions for lesser offenses at much higher rates than their counterparts (Skiba & Knesting, 2001). When aversive strategies are the primary means of controlling behavior, several negative side effects occur including increases in anti-social behavior, more coercive interactions between adults and students, and a decrease in academic achievement appropriate social behaviors (McEvoy & Welker, 2000; Skiba & Peterson, 2000).

Positive Behavior Interventions and Support (PBIS) is an alternative response to challenging behavior that is proactive, preventative, and able to facilitate effective change in schools and individual students (Sugai & Horner, 2002). In order to improve educational outcomes for all students, federal initiatives stress the importance of evidence-based practices. As described in the widely-referenced document, “Is School-wide PBIS and Evidence-Based Practice” (OSEP Technical Assistance Center on PBIS, 2009), evidence-based is demonstration of at least two peer-reviewed randomized control trial research studies that document experimental control. To meet this standard the practice/procedure must be operationally defined, formal measures of fidelity must be applied, formal outcome measures must be
specified and monitored, and these elements must be used within a randomized control trial group research design (Gersten, et al., 2005). Positive Behavioral Support (PBS) and Positive Behavioral Interventions and Support (PBIS) have both appeared in the literature and for the purposes of this research are considered synonymous. PBIS as the preferred language will be used throughout this report. Appendix A contains a consolidated list of acronyms and abbreviations to assist the reader.

The No Child Left Behind (NCLB) federal legislation has created a new way to look at academic progress. Schools have begun to use research-based practices to teach students reading, writing, and mathematics. In addition, NCLB requires that states identify “persistently dangerous schools.” PBIS implementation is a way to impact the learning environments in all schools in order to support high academic performance and to reduce behavior problems. As Sugai and Horner (2008) note, “The academic and behavior link is clear: good instruction is one of our best behavior management tools, and positive and preventative behavior management are some of our best instructional support strategies.”

As with all initiatives, researchers and educators are charged with identifying evidence-based practices. In a report by the What Works Clearinghouse (Institute of Education Sciences, 2008) Epstein, et al. identified strong evidence for the following practices: defining and teaching behavioral expectations, modifying the physical environment to discourage problem behavior, individualizing instruction to promote high rates of engagement, and teaching and reinforcing prosocial skills. These practices are all key components of Positive Behavior Interventions and Support.
PBIS is defined by the Office of Special Education Programs (OSEP), Technical Assistance Center on Positive Behavioral Interventions and Supports (2009) as “an application of a behaviorally based systems approach to enhance the capacity of schools, families, and communities to design effective environments that improve the fit or link between research validated practices and the environment in which teaching and learning occur.” Schools strive to implement PBIS as an effective and proactive process for improving social competence and academic achievement for all students.

PBIS is focused on the design of environments that promote desired behaviors and minimize the development and support of problem behaviors. Applied behavior analysis (ABA; Baer, Wolf, & Risley, 1968) is the conceptual foundation for these empirically proven intervention practices. ABA is grounded in the assumption that human behavior can change and provides a conceptually powerful operant model for validating support to address the unique needs of individuals with problem behavior (Dunlap, Sailor, Horner, & Sugai, 2008).

One of our nation’s top priorities is to keep schools safe by providing a place where students can learn and teachers can teach free from threats of harm. Although positive student outcomes are the intended benefits of implementing universal PBIS, to achieve those desired student outcomes the practices and processes must be adopted and implemented with fidelity. One of the greatest challenges to achieving wide spread educational reform is the actual implementation of educational programs and practices. Fixsen and colleagues (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005) reviewed the literature on successful program implementation in attempting to identify variables related to effective installation/adoption of educational programs and their subsequent sustainability. According to Fixsen et al., (2005) the
stages of the implementation process are adoption, program installation, initial implementation, full operation, innovation, and sustainability. Sustainability is the “durable, long-term implementation of a practice at a level of fidelity that continues to produce valued outcomes” (McIntosh, Horner, & Sugai, 2009). The critical mechanism by which a practice is sustained is fidelity of implementation (McIntosh, Filter, Bennett, Ryan, & Sugai, 2010). With universal PBIS there are two tools that are used to measure implementation with fidelity: School-wide Evaluation Tool (SET) and Benchmarks of Quality (BOQ). The main difference between these two instruments is that the SET is done by an external evaluator and the BOQ is a self-assessment. Both produce scores from 0 to 100, with 0 implying low fidelity and 100 reflecting high fidelity.

The first of the evaluation tools for implementation of school wide Positive Behavior Interventions and Support is the Benchmarks of Quality (BOQ; Kincaid, Childs, & George, 2005). This instrument is a rating scale containing 53 items, organized into 10 principal components or critical elements of PBIS implementation. This self assessment is completed by school teams on a yearly basis to assess how they score with regard to developing and implementing school wide PBIS. These scores guide each school’s action plan for the following year.

The second of the fidelity of implementation measures for PBIS is the School-wide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001). The SET involves a 2 to 3 hour site visit to a school by an external evaluator. During the site visit, there are observations and interviews with administration, staff, and students. Seven primary feature of universal or school-wide PBIS are measured including the extent to which schools have developed and implemented rules, consistently taught and reinforced rules and expectations, consistently
responded to behavior problems, and used teamwork, leadership, and data for decision-making. SET outcomes have been found to be reliable and valid measures of implementation at the universal level and are closely linked to other indicators of school climate and safety (Horner et al., 2004).

**Context of the Problem**

Positive Behavior Interventions and Support was first implemented in Louisiana in 1993 to serve the students with more severe behavior issues. To serve all students by implementing universal PBIS, trainings began in 2002 for school teams. As of April 2010, trainings have continued for the universal level, and at the current time 979 of the approximately 1600 schools across the state have been trained in universal PBIS. For the past 3 years, data have been collected on over 1000 schools to determine the implementation level of universal PBIS. Collectively, over 1000 schools have submitted a Benchmarks of Quality (BOQ) in the spring with the state average of 72%. A BOQ score of 70% or over is considered an indicator of high implementation. Over 1017 schools have been evaluated using the School-wide Evaluation Tool, and 548 schools have qualified as a high implementing school by scoring an 80/80 on the SET. Although many schools across the state have been trained and are considered to be implementing universal PBIS, there has been limited investigation of the relationship between levels of implementation and student and school outcomes.

**Purpose of the Study**

The current study examines the relationship between universal implementation of PBIS and student achievement on the LEAP and iLEAP examinations administered as part of the Louisiana Educational Assessment Program. To address this question, it is necessary to take
into consideration individual as well as school level factors that may act to facilitate or impede student achievement. Multilevel statistical models are ideally suited for research problems of this nature and are the approach taken for this study (Luke, 2004). In particular, as statistical controls at the student level, we include measures of prior academic achievement. At the school level, the control variables include average student achievement and percentage of free/reduced lunch. The variables considered in the model can be grouped into characteristics of (a) the student population, (b) instructional staff, and (c) the school organization. The research questions explored in this investigation are:

1) Is the degree of implementation of the universal level of PBIS related to changes in rates of discipline actions at the school level?

2) Are measures of implementation (BOQ, SET) of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP?

3) Do identified student/school characteristics contribute to any difference in academic performance?

The contributions for this study will extend the use of SETs and BOQs as measures of school level implementation. This study partially replicates the analysis of Horner, Sugai, Eber, and Lewandowski (2004) and Simonsen et al. (2012) and of a statewide evaluation in Illinois of the impact of PBIS. A more stringent criterion of suspensions/expulsions as opposed to office discipline referrals and practical consideration will be considered. This will become one of the largest analyses of students and schools to date and provides the first use of a multi-level
analysis in examining the impact of PBIS on academic behavior thus possibly identifying mediating or contributing variables to successful implementation.
CHAPTER 2: LITERATURE REVIEW

Since its emergence in the mid-1980s, PBIS has developed rapidly as a broad and multifaceted approach for addressing difficulties in behavioral adaptation and for encompassing interventions addressed not only to enhancing individual as well as collective lifestyles (Sailor, et al., 2009). From its inception in the disability rights movement and Applied Behavior Analysis foundations, PBIS evolved from the requirement of P.L. 105-17, titled the Individuals with Disabilities Education Act (IDEA) of 1997.

Early research using PBIS was conducted with individuals with severe cognitive and developmental disabilities and targeted such behaviors as aggression, self-injury, and property destruction; however, with the broad goal of creating safe schools, the use of PBIS expanded to include all students (Safran & Oswald, 2003). PBIS has been defined as an application of a behaviorally-based systems approach to enhance the capacity of schools, families, and communities to design effective environments that improve the fit or link between research-validated practices and the environments in which teaching and learning occurs (www.pbis.org). Attention is focused on creating and sustaining primary (school-wide), secondary (classroom), and tertiary (individual) systems of support that improve lifestyle results for all children and youth by making problem behavior less effective, efficient, and relevant, and desired behavior more functional. PBIS schools organize their evidence-based behavioral practices and systems into an integrated continuum in which students experience supports based on their behavioral responsiveness to intervention. PBIS represents a three-tiered prevention logic requires that all students receive supports at the universal or primary tier. If the behavior of some students is not responsive, more intensive behavioral supports are
provided, in the form of a group contingency (selected or secondary tier) or a highly individualized plan (intensive or tertiary tier).

On a school-wide level, PBIS relies on accurate and reliable discipline referral data to understand the behaviors occurring across campus. An analysis of the data allows a school team to identify the problem areas, brainstorm interventions such as where and what to teach, reward the students exhibiting the expected behavior, and communicate findings to the staff, students, and families. School-wide Positive Behavior Interventions and Support establishes and reinforces clear expectations for behavior. It is a team-based system that requires participation of teachers, administrators, families, and students and a common approach to discipline that is proactive, instructional, and outcome-based.

Data guide decision-making with teams using information from the entire school and the whole school day. Improving student academic and behavior outcomes is about ensuring that all students have access to the most effective and accurately implemented instructional and behavioral practices and interventions possible. Universal PBIS provides an operational framework for achieving these outcomes. More importantly, PBIS is not a curriculum, intervention, or practice, but it is a decision making framework that guides selection, integration, and implementation of the best evidence-based academic and behavioral practices for improving important academic and behavior outcomes for all students. In general, schoolwide PBIS emphasizes four integrated elements:

(a) data for decision making,

(b) measurable outcomes supported and evaluated by data,

(c) practices with evidence that these outcomes are achievable, and
(d) systems that efficiently and effectively support implementation of these practices.

PBIS programs provide a comprehensive process for teaching socially appropriate behaviors in order to optimize academic achievement for each student. PBIS provides a positive and effective alternative to the traditional methods of discipline, and its methods are research-based to significantly reduce the occurrence of problem behaviors in the school, resulting in a more positive school climate and increased academic performance. PBIS is consistent with the Individuals with Disabilities Education Act, which advocates the use of positive behavior interventions and school-based disciplinary strategies that reduce or eliminate the need to use suspension and expulsion as disciplinary options.

**History of PBIS**

In 1997, Congress amended the Individuals with Disabilities Education Act (IDEA); positive behavioral support has held a unique place in special education law. PBIS is the only approach to addressing behavior that is specifically mentioned in the law. The emphasis in the law states that utilizing functional behavior assessments and positive approaches to encourage good behavior remains in the current version of the law as amended in 2004. Congress’ reasons for encouraging the use of PBIS is clear and stem from (a) the historic exclusion of individuals with disabilities based on unaddressed behavior and (b) the strong evidence base supporting the use of PBIS. Congress recognized in 1997 and 2004 the need for schools to use evidence-based approaches to proactively address the behavioral needs of students with disabilities. Congress explicitly recognized the potential of PBIS to prevent exclusion of disabled students and the improvement of educational results.
Systems of PBIS

According to Reynolds, Irwin, and Algozzine (2009), PBIS is a systems approach or process, not a specific curriculum. The goal is to help educate all students, especially those with challenging behaviors. Because of the emphasis on continuous, data-based improvement, each school individualizes its implementation of PBIS. The adoption and sustained use of effective leveled-practices is central to PBIS and there is a strong emphasis on systematically teaching behavior using effective instructional methods, approaches, and practices. There is support for teaching and encouraging pro-social behavior to small groups of students, as well as for systematically teaching and implementing behavior interventions for the most difficult students. Positive Behavior Interventions and Support (PBIS) is based on a problem-solving model and aims to prevent inappropriate behavior through teaching and reinforcing appropriate behaviors (OSEP Technical Assistance Center on Positive Behavioral Interventions & Supports, 2009).

PBIS offers a range of interventions that are systematically applied to students based on their demonstrated level of need, and addresses the role of the environment as it applies to development and improvement of behavior problems. Both Response to Intervention (RTI) and PBIS are grounded in differentiated instruction. Each approach has critical factors and components to be in place at the universal (Tier 1), targeted group (Tier 2), and individual (Tier 3) levels.

Benchmarks of Quality (BOQ)

The BOQ is a 53 item rating scale used to assess the degree of fidelity with which a school is implementing universal PBIS. This is a self-assessment completed by school PBIS team
members with the help of an outside facilitator. The 53 items correspond to the 10 subscales of the instrument: PBIS Team, Faculty Commitment, Effective Discipline Procedures, Data Entry, Expectations and Rules, Reward System, Lesson Plans, Implementation Plans, Crisis Plans, and Evaluation. The total possible score for the BOQ is 100, and this score is derived from three to eight items in each of the 10 subscales. Each item has a maximum value between 1 and 3.

Psychometric properties of the instrument for reliability are an overall Cronbach’s alpha of 0.96. Test-retest reliability was computed with Pearson product-moment correlations of 0.94 ($p<0.01$). Inter-rater reliability indicated a high correlation of 0.87 ($p<0.01$; Kincaid, Childs, & George, 2005).

Three components of the Benchmarks of Quality include the Team Member Rating form, scoring form, and scoring guide. The Team Member Rating Form is to be completed by team members independently and returned to coach at the school. The Scoring Form is completed by the coach using a Scoring Guide that is used for reporting back to team. The Scoring Guide describes procedure for completing BOQ and includes a rubric for scoring each item. The Benchmarks of Quality will include a summary of team members’ perceptions of PBIS implementation (scored: ++ in place, + needs improvement, and - not in place). This objective assessment of school’s implementation is based on criteria described in a rubric (100 point scale). The comparison between the above factors will encourage discussion of strengths and weaknesses and provides ideas for action planning.

**School-wide Evaluation Tool**

Another evaluation tool to determine the implementation of school-wide PBIS is the School-wide Evaluation Tool (SET). It was developed by the National Positive Behavior
Interventions and Supports Center and is widely used in PBIS evaluation efforts (Horner, et al., 2004). The SET is designed to assess and evaluate the critical features of school-wide Positive Behavior Interventions and Support across each academic school year. External observers gather information necessary for this assessment tool through multiple sources including review of permanent products, observations, and staff and student interviews or surveys. The SET is a research instrument for determining the extent to which a school is implementing school-wide Positive Behavior Interventions and Support. The SET evaluates a total of twenty-eight research questions across the seven feature areas. The features include (a) expectations defined, (b) behavioral expectations taught, (c) acknowledgement procedures, (d) correction procedures, (e) monitoring and evaluation, (f) management, and (g) district level support.

Information necessary for this evaluation tool is gathered through multiple sources including a review of permanent products (including discipline handbook, school improvement plan for safety related goal, instructional materials, meeting minutes), observations, and brief staff and student interviews.

The scoring guide is used for scoring the level of implementation on each of the twenty-eight research questions. It is organized by feature area and is formatted to provide the research question and the criteria for scoring each question. Each SET research question has two possible points. Using the established criteria for each question, the SET evaluator determines a 0, 1, or 2 point score for each question. The SET results provide a summary score that is used (a) to determine annual goals for school-wide effective behavior support, (b) to evaluate on-going efforts toward school-wide behavior support, (c) to design and revise procedures as needed, and (d) to compare annual accomplishments toward school-wide
effective behavior support. The SET is useful for assessing and evaluating the critical features of school-wide effective behavior support. The SET has strong psychometric properties of internal consistency (Cronbach’s alpha= 0.96), strong test-retest reliability, and high interobserver reliability (range= 98.4%-100% across 17 schools; Horner et al., 2004). To indicate a high level of universal PBIS implementation, an 80% overall and 80% on Subscale 2 or teaching component (“80/80”) is widely recognized as the criterion for full universal PBIS implementation (Horner, Sugai, Todd, & Lewis-Palmer, 2005; Horner et al, 2009).

**Implementation and Academic Achievement**

Horner, Sugai, Todd, and Lewis-Palmer (2005) expressed that the level or fidelity of implementation is most critical and often ignored aspect of interventions. The overall issue is whether schools/teachers implement interventions as planned, with sufficient intensity and integrity to reproduce expected outcomes. Fixsen et al. (2005) states that even the most effective intervention will not produce positive effects if it is not implemented. Therefore, assessments of the performance or results of the intervention are a critical component of implementation. Wilder, Atwell, and Wine (2006) show that different levels of integrity result in different levels of outcomes. Northup, Fisher, Kahng, Harrel, and Kurtz (1997) additionally suggest high integrity followed by declines in integrity has limited impact on outcomes.

Kovaleski, Gickling, Morrow, and Swank (1999) evaluated high vs. low implementation of Instructional Support Teams (IST). The results indicated school-wide organizational change, where students benefited from IST processes only when implemented with high fidelity. They also showed that implementing with low fidelity resulted in no better outcomes for students than control group not exposed to IST processes and that having structures in place was not
sufficient to assure high fidelity. Horner et al. (2005) studied the effect of high fidelity vs. low fidelity of implementation of PBIS on office discipline referrals. They found schools that implemented with high fidelity had 25% fewer office referrals for major rule violations than schools that did not meet fidelity criterion. In this study fidelity measures were taken two times per year.

**Academic Achievement and PBIS**

The logic that follows this research question is that through the framework of Positive Behavior Interventions and Support, classroom management and curriculum variables would be adapted so academic tasks become less aversive. Reduction in office discipline referrals would result in more minutes spent in academic instruction, the minutes spent in academic instruction would be more effective, there would be less peer support for academic failure, and there would be an increase in the structured prompts, contingent feedback and support for academic behavior. With these conditions in place, we hypothesize that a school could affect the academic gains of students (Putnam, Horner & Algozzine, 2006).

Several studies have documented patterns in office discipline referrals within schools, relevant to this study as if students are not in class they cannot benefit from instruction. Kaufman et al. (2010) found gender differences in the rate and type of referrals with significantly more referrals for boys’ delinquent and aggressive behavior; as well as significantly more referrals for African American students. Another study with administrative decisions from 1,510 schools nationwide in 2005-2006 showed that most ODRs were generated from classrooms, and peer-directed for elementary students, and adult-directed for middle school students (Spaulding et al., 2010).
Research has consistently shown that the amount of instructional time is highly correlated with student achievement (Brophy, 1988). As it follows if acceptable instruction is in place, then improving the behavioral climate of the school will allow that instruction to be more effective. When schools are unsafe and chaotic, academic learning is less likely to occur (Kellam, Mayer, Rebok, & Hawkins, 1998). Putnam suggests that class-wide behavior support increased the time students receive academic instruction. (Putnam, Handler, Rey and O’Leary-Zonarich, 2002). The implementation of PBIS itself does not lead to direct academic gains; it can only mediate the effectiveness of the existing academic instruction.

Along those same lines, if students are excessively absent or being suspended or expelled from school, they are not able to benefit from instruction because they are not there. Schachter (2010) found from very large school districts that by implementing PBIS they were able to decrease suspensions and increase academic achievement. In Denver, out of school suspensions went from 14,000 in 2002-2003 to about 8,000 in 2009. In the Los Angeles Unified School District, one school in central L.A. grew an average of 55 points on the California Standardized Test the first year of implementation (Schachter, 2010).

There is a documented connection between low academic skills and problem behavior, which may be evident as early as kindergarten but grows over time as students move from elementary to secondary school. Fleming, Harachi, Cortes, Abbott, and Catalano (2004) examined 783 students from 10 schools in the Pacific Northwest. Latent growth curve models for reading and teacher report of attention problems from grades 3 to 6 showed statistically significant heterogeneity in initial level and change. Morrison, Anthony, Storino, and Dillon (2001) studied characteristics for students referred to an in-school suspension program for
disciplinary infractions in middle school were analyzed. Almost half of the students referred had previous office referrals, and one-third had previous suspensions. The students referred did not have attendance problems as reported by the assistant principals, but they did have a low “C” average in their grades. Those with previous office referrals averaged below a “C” grade point average (Morrison et al., 2001). Nelson, Benner, Lane, and Smith (2004) indicated that students with emotional and behavioral disorders have deficits in all content areas of academic achievement. There were no differences found based on gender. Another purpose of the study was to identify types of problem behavior that are related to academic achievement. The results suggest that externalizing behaviors (i.e. attention, aggression, delinquency) of the students in the sample were the only construct that contributed to the overall fit-of-the-model related to their academic achievement in all content areas (Nelson et al., 2004). There is evidence that problems in reading or behavior can predict future problems in other areas. Poor academic skills early in school predict a wide range of behavior problems, because students who have difficulty with reading may find problem behavior as an effective means of escaping or avoiding reading activities (McIntosh, Horner, Chard, Dickey, & Braun, 2008). Students may engage in problem behaviors because the academic activity may be too difficult, too easy, or not relevant to student needs or interests.

McIntosh, Horner, Chard, Boland, and Good (2006) found that kindergarteners with phonological awareness skills, as measured through the Dynamic Indicators of Basic Early Literacy Skills Phoneme Segmentation Fluency subtest (Good & Kaminski, 2002), that indicated low risk for reading problems (at least 35 sounds) had an 18% chance of receiving two or more office discipline referrals (ODRs) in 5th grade. Students scoring in the some risk range (between
10 and 35 sounds) had a 25% chance of having multiple ODRs. Students with scores in the at-risk range (below 10 points) had a 33% chance of multiple ODRs in 5th grade. In summary, the authors concluded the two most powerful predictors of ODRs for fifth grade students were a spring DIBELS score in kindergarten and fourth grade ODRs.

The overall picture provided by McIntosh’s results in one in which children enter kindergarten with varying reading skills. If, however, they do not respond to literacy instruction during kindergarten, and fall behind, a negative spiral of achievement and behavior becomes more likely. As the student’s literacy skills do not keep pace with those of peers, academic tasks become more aversive, and problem behaviors that lead to escape from these tasks become more likely (Putnam, et al., 2006). Moreover, a replication study indicated that students who entered school with phonological awareness deficits but responded to kindergarten reading instruction were at dramatically decreased risk for future problem behavior (McIntosh, Sadler, & Brown, 2009).

In a more recent study, McIntosh, Sadler and Brown (2012) used a longitudinal data set of 473 students, including Dynamic Indicators of Basic Early Literacy Skills measures at the start, middle, and end of kindergarten and office discipline referrals in Grade 5. This was used to determine whether reading skills at school entry or change in reading skills over the course of kindergarten were more predictive of chronic problem behavior in Grade 5. Results of logistic regression analyses found that low initial phonological awareness predicted problem behavior, but also including skill growth during kindergarten in the model resulted in significantly improved and more accurate prediction.
Luiselli, Putnam, Handler, and Feinberg (2005) concluded that an urban elementary school in the Midwest showed reading comprehension and mathematics percentile ranks on standardized tests improved from the first (pre-intervention) to the second (intervention) test dates, increasing 18 and 25 percentage points respectively (Luiselli et al., 2005). This study illustrated that schools implementing School-wide PBIS have decreased problem behaviors, increased time spent in academic instruction, and are associated with improved academic outcomes on standardized test scores.

In a randomized, wait-list controlled study assessing PBIS in elementary schools in 2 states over a 3 year period, showed that training and technical assistance were functionally related to improved implementation of universal PBIS. PBIS implementation was also associated with the percentage of third graders meeting the state reading standard. There were statistically significant differences between T1 and T2 for the treatment group (0.056, t (57)=2.75, p=0.008) (Horner, Sugai, Smolkowski, Eber, Nakasato, Todd, & Esperanza, 2009).

Lassen et al. (2006), examined school-wide PBIS implementation in an urban, inner-city middle school in the Midwest over a 3-year period. Data on ODRs, suspensions, standardized test scores, and treatment fidelity were gathered and analyzed. Results demonstrated significant reductions in ODRs and suspensions and increases in standardized math and reading scores. Additionally, regression analyses suggested a significant relationship between student problem behavior and academic performance, especially in math. Fidelity of implementation of PBIS procedures was significantly correlated with reductions in problem behavior. The authors also raise the issue that fidelity of implementation has not been included as a measure in many of the studies dealing with PBIS and outcome measures, although it was included in this study.
The authors further add that it is hard to answer the question, “Did the intervention occur as intended?” without a measure of treatment fidelity.

As Simonson, et al. (2012) note, “After performing a search of the PsychInfo and ERIC databases for the descriptors of ‘school wide positive behavior support’ and ‘fidelity,’ it appeared that no published peer-reviewed studies have systematically examined the effects of implementing SWPBS with (and without) fidelity over time.”

**State-wide PBIS and Academic Achievement**

Several studies have been published documenting the effectiveness of state-wide PBIS initiatives. In the evaluation of North Carolina School-wide PBIS (Irwin & Algozzine, 2005), with 264 elementary schools in North Carolina that have adopted PBIS there were able to show promising findings. After one year of implementation, there were statistically significant increases in fifth grade promotion and in composite performance defined as the percent of students performing at grade level or higher on the end of grade or end of course tests.

In Hawaii, one school with a large number of disciplinary incidents reviewed its incident data and learned that 56 percent of its behavioral incidents were attributed not attending class and 15 percent to tardies). School staff used this information to adopt efforts to improve class attendance and punctuality. The school’s prevention efforts focused on the strong correlations between academic achievement and attendance, punctuality, and appropriate classroom behavior (Nakasato, 2000).

In the New Hampshire state PBIS Initiative there are two sets of finding that show the link between universal implementation and outcomes. With the first publication, the main finding was to determine the level of implementation. Muscott, Mann, Benjamin, and Gately
(2004) showed that 54% or 15 of the 28 schools achieved a high level of implementation status as defined by 80/80 on the SET. In a follow up study in New Hampshire (Muscott, Mann & LeBrun, 2008) there were 28 schools (as above mentioned) implementing PBIS in first cohort (1 pre-elementary, 13 elementary, 6 middle school, 4 high school, and 4 mixed level schools). There was shown to be a reduction of 6,010 ODRs and 1,032 suspensions. These reductions recovered 1,701 days of student learning. In New Hampshire, PBIS Implementation was associated with academic gains in math occurred in 16/22 (73%) of schools that achieved 80% or better on the SET and academic gains equal increase in percent of students achieving basic or above on state math test.

Another statewide evaluation in Maryland were able to show substantial gains in the outcome measure of suspensions by showing a statistically significantly reduction with 62 schools ($p= 0.03$; Barrett, Bradshaw, & Lewis-Palmer, 2008). In Iowa’s statewide PBIS initiative 75% (18 out of 24) showed a 42% average rate of decrease of ODRs per day per 100 students across a 2 year period (Mass-Galloway et al., 2008).

Horner et al. (2004) in reference to the Illinois PBIS research, at the universal PBIS level for schools achieving 80/80 on the SET, 62% of 3rd grade students met the Illinois State Achievement Test Reading Standard (n=52). For schools not achieving 80/80 on the SET, only 47% of 3rd grade students met the Illinois State Achievement Reading Test Standard (n=69).

Significant improvements in learning were reported from a comprehensive school-wide PBIS program. Working with 7 elementary schools in a school district, compared to the district’s remaining 28 elementary schools in the Pacific Northwest, there were found to be significant positive differences in 4th grade reading, language arts, spelling, science, and social
studies achievement on the California Test of Basic Achievement, as well as significant positive differences in the reading and mathematics scores on the Washington Assessment of Student Learning. It is also an important note that target students who exhibited problem behaviors improved their academic performance, whereas the criterion students’ academic performance remained relatively stable. There was a consistent decline in suspensions, removals, and office referrals for the participating schools; in contrast there were increases in these administrative discipline actions for the non-participating schools (Nelson, Martella, Marchand-Martella, 2002).

The use of office discipline referrals (ODRs) and suspensions and expulsions are typically used as outcomes measures of PBIS. ODRs are the standard outcome measure showing the most immediate changes with implementation of PBIS. This is the ideal measure to determine if a school is making an effect with implementation at the school level. Although the most sensitive measure being ODRs, many states may not have this level of data collection and that from a systems level, suspensions and expulsions are the data that are collected and recognized. This is particularly true in Louisiana where the state Department collects this data of suspension and expulsion and not ODRs. It should also be mentioned that suspensions and expulsions in Louisiana are used as part of a school’s accountability to the State Department of Education.

Current research is encouraging, but remains descriptive in nature and does not have the experimental control needed to confirm a relationship between school-wide PBIS and improved academic performance. Decreases in problem behavior and increases in time for learning and teaching are all important and necessary outcomes of any school-wide discipline
system; however in the current era of academic accountability, they are insufficient if not paired with academic achievement gains.

**PBIS in Louisiana**

To lay the context of this research, in Louisiana, as well as many other states, Positive Behavior Interventions and Support has been implemented as the approach to help schools in the area of behavior. The information below from the Request for Proposal provides some baseline data for the current project.

There are a high number of Louisiana students who are removed from classroom instruction as a disciplinary action. In the 2004-2005 school year there were 79,133 In-School Suspensions (ISS), 92,193 Out-of-School Suspensions (OSS), 4,143 In-School Expulsions (ISE), and 3,933 Out-of-School Expulsions (OSE). While some expulsions are necessary for the protection of the general student body, many students are removed from school for repeated minor (i.e., non-life threatening) violations. The use of removals as a disciplinary action in Louisiana has been consistently increasing over the past 5 years. From 2000 through 2005, there have been average annual increases in the percentage of students receiving In-School Suspensions, Out-of-School Suspensions, In-School Expulsions and Out-of-School Expulsions of 4.9%, 5.0%, 15.7% and 6.0%, respectively. These increases in the use of exclusionary practices indicate that removal has not been effective in addressing (i.e., decreasing) challenging behaviors in our schools. It is necessary for schools to adopt more proactive, educational, data-driven approaches that are effective in reducing the need for reactionary, exclusion-based approaches to rule violations and result in increased instructional time for all students (Louisiana Department of Education, 2006).
In Louisiana, there is an inverse relationship between suspension rates and student performance. The 2004-2005 Louisiana State Education Progress Report indicates that on average higher performing schools had 10% fewer student suspensions than lower performing schools. According to some educators in Louisiana, many of the students removed for disciplinary infractions are the same students who need more, not less, instructional time (Louisiana Department of Education, 2006). In addition, there is a historical pattern of disproportional representation across ethnic and gender subgroups in the performance of Louisiana state academic standards, the use of exclusionary disciplinary practices, and the identification for disabilities (specifically emotional/behavioral disorders and specific learning disorders). Specifically, African-American males tend to have lower academic performance, are more likely to be removed for disciplinary reasons, and more likely to be identified as having an emotional and/or learning disability than other subgroups (Louisiana Department of Education, 2006).

In 2003, the Louisiana Legislature passed the Juvenile Justice Reform Act 1225, which provides direction to the Board of Elementary & Secondary Education (BESE). The Education/Juvenile Justice Partnership Act legislated that the Board of Elementary and Secondary Education (BESE) would formulate, develop and recommend a Model Master Plan for improving behavior and discipline within schools that includes the utilization of positive behavioral supports and other effective disciplinary tools. In addition, each city, parish, and other local public school board should be responsible for the development of school master plans for supporting student behavior and discipline based upon the model master plan developed and approved by BESE (Louisiana Department of Education, 2006).
The current study is to examine if implementation of the universal level of PBIS is related to student achievement on the LEAP and iLEAP examinations administered as part of the Louisiana Educational Assessment Program. A second purpose is to examine whether identified student/ school characteristics contribute to any difference in academic performance. To address this question, it is necessary to take into consideration individual as well as school level factors that may act to facilitate or impede student achievement. Multilevel statistical models are suited for research problems with the structure of students nested within schools and will be the approach taken for this study. In particular, as statistical controls at the student level, we will include measures of a student’s prior academic achievement. At the school level, the control variables will include average student achievement, percentage of free/reduced lunch, and the school PBIS implementation score. The research questions explored in this investigation are:

1) Is the degree of implementation of the universal level of PBIS related to changes in rates of discipline actions at the school level?

2) Are measures of implementation (BOQ, SET) of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP?

3) Do identified student/ school characteristics contribute to any difference in academic performance?
CHAPTER 3: METHODOLOGY

The current chapter provides information about data collection methods and statistical procedures involved in the present study. It begins with a description of the sample and population. A discussion of the variables used and their operational definitions are discussed. Finally, the statistical models and their analysis strategies are presented.

A multilevel statistical model will be used to address the primary research question of this study. Multilevel models are ideally suited to studies of this nature because they permit explicit modeling of within and between school processes. At the within-school level, we specify a student level model in which student achievement on the LEAP examinations is a function of (a) prior achievement.

At the school level we focus on explaining variation in the intercepts of the student model. These intercepts are grand-mean centered so as to permit interpretation as average student achievement, adjusted for prior academic achievement. The predictors in the school level model include (a) level of PBIS implementation, (b) average student achievement, and (c) percent of free and reduced lunch.

Multilevel Model

Multilevel models are rapidly becoming a cornerstone of modern statistical analysis in the behavioral, social, and health sciences (Luke, 2004). For example, in educational research it may be of interest to compare the achievement of students in schools in which teachers utilize team-based instruction to the achievement of students in schools in which instructional teams are not used. It would be a mistake to analyze the data as though the students were randomly assigned to the different teaching models. A more accurate representation of these data would
reflect the fact that schools are likely to have a unique effect on student outcomes, irrespective of the import of different teaching models. In other words, the performance of students within the same school will be correlated and these correlations must be represented in the analysis for correct inference to be drawn from the study. According to Bauer and Curran (2009), the advantages of multi-level models are twofold. First, unlike more traditional statistical models such as ANOVA or regression, multilevel models provide accurate results when applied to data sets consisting of hierarchically clustered observations (e.g., individuals within groups) or repeated measures. Secondly, when used effectively, multilevel models can provide new insights into phenomena that were previously poorly understood, for instance by helping to identify the level at which predictors exert their effects.

Multilevel models can be used to model change over time in a variable of interest. The variable of interest in this study is academic achievement as represented with state–level testing scores. The statistical theory behind multilevel modeling has evolved over the years. When confronted with difficult conceptual problems, many researchers view the data as nested data structures. The problem with this structure is that it violates the assumption of independence required by traditional statistical analysis such as ANOVA and ordinary least-squares (OLS) multiple regression. Peugh (2010) states, for example, that response variable scores of students in the same school are likely to be more correlated than for students within different schools because they share the same learning environment. These independence violations tend to make multilevel modeling a necessity because traditional analysis can produce excessive Type 1 errors and biased parameter estimates.
According to Luke (2004), one alternative approach to doing a multi-level model analysis has been to disaggregate group-level information to the individual level so that all predictors in a multiple regression model are tied to the individual unit of analysis. This leads to at least two problems. First, all of the un-modeled contextual information ends up pooled into the single individual error term of the model (Duncan, Jones, & Moon, 1998). This is problematic because individuals belonging to the same school will presumably have correlated errors, which violates one of the basic assumptions of multiple regression. The second problem is that by ignoring context, the model assumes that the regression coefficients apply equally to all contexts, “thus propagating the notion that processes work out in the same way in different contexts” (Duncan et al., 1998, p. 98).

One partial solution to these statistical problems is to include an effect in the model that corresponds to the grouping of the individuals. This leads to an ANOVA or ANCOVA approach to modeling. Unfortunately, there are still a number of issues with this approach. First, in the case where there are many groups, these models will have many more parameters, resulting in greatly reduced power and parsimony. Secondly, these group parameters are often treated as fixed effects, which ignores the random variability associated with group-level characteristics (Luke, 2004). Due to the modeling of the ANCOVA, the context specific characteristics cannot be modeled directly. The ANCOVA is able to answer the research question, “Do schools differ?” The more important question of “Why schools differ?” cannot be answered by this approach (Kreft & Leeuw, 1998). Finally, ANOVA methods are not very flexible in handling missing data or greatly unbalanced designs.
Multilevel models have been used in education research to estimate separately the variance between students within the same school, and the variance between schools. The structure of this model fits the model of this research question. As students and nested within schools, a multilevel approach to answering the research questions will be best suited.

Sample and Population

The sample used in this study includes students and schools in the state of Louisiana that have been trained in School wide Positive Behavior Interventions and Support. Some schools have participated for several years and are at full implementation, while others are relatively new to the program and have not achieved full implementation, as discussed in the following sections. According to research, a highly implementing school has scored an 80% total score and 80% on the teaching component on the SET or a 70% on the BOQ (Horner et al. 2004; Kincaid et al., 2005). This is the standard that is accepted nationally and used in many of the other state’s evaluation plans. The entirety of elementary schools in Louisiana that participated in PBIS during the 2007-2008 and 2008-2009 school years will serve as the targeted population. Because archive data are available on these schools and their students, all schools in this targeted population will be included in the quantitative analysis described below. Students tested in these schools as part of the Louisiana Educational Assessment Program during spring 2009 are also included in this study.

Schools in Louisiana that are implementing PBIS have chosen to become trained and participate in the project. Because participation in the project is voluntary, the targeted population is not necessarily representative of the larger population of schools in the state, although 979 of the state’s approximately 1600 schools have been trained.
Measures/Variables

The primary objective of this study is to determine if fully implementing the PBIS universal program impacts student achievement. To accomplish this objective, this study focuses on the relationship between school level participation in the program, defined as level of implementation, and student performance on the standardized tests administered as part of the Louisiana Educational Assessment Program. Because many factors can impact student outcomes on these examinations, both student and school level control variables will be included in the study. Student level control variables capture the features of the student that are likely to influence outcomes. In this study these included are measures of prior academic achievement. School level control variables include features of a school that influence outcomes. These include average student achievement, PBIS implementation level, and percent of students received free or reduced lunch at a school. Outcome data are long range school-level data that result from the interventions implemented by an effective Positive Behavior Interventions and Support process. Process data reflect a schools level of implementation of Positive Behavior Interventions and Support.

These data were obtained from the Louisiana Department of Education (LDOE) through the Principal Investigator. In no case is student identification made part of the analysis. Student identifying information remained with the Louisiana Department of Education, and a unique student identifier was given by LDOE. All other data used in this analysis are public information and available online or through the LA-PBIS project.
Operational Definitions

In this section the operational definitions of the variables used throughout the study are presented.

Achievement Scores

LEAP-21 and LEAP scores will be used as the determination of academic achievement for 3rd grade and 4th grade.

Louisiana Education Assessment Program for the 21st Century (LEAP 21)

According to the Louisiana Department of Education (2009), LEAP 21 is the state sanctioned criterion-referenced assessments used to monitor how well students have learned the state content standards up to fourth and eighth grades. The purpose of LEAP 21 was therefore, to ensure that fourth and eighth grade students had the knowledge to pass onto the next level of education (i.e., middle school and high school). To determine this, students are given achievement ratings; advanced, mastery, basic, approaching basic and unsatisfactory. In order for students to move to the next grade, they must have and achievement rating of approaching basic or above.

Integrated Louisiana Educational Assessment Program (iLEAP)

According to the Louisiana Department of Education (2009), the iLEAP is a criterion-referenced test administered to students in third, fifth, sixth, seventh, and ninth grade in the state of Louisiana. The iLEAP assigns students into one of five categories based on their performance: unsatisfactory, approaching basic, basic, mastery, and advanced. The ELA test consists of four subtests, reading, language, writing, and using information resources. The subtests include 78 multiple-choice items and a writing prompt. The reliability of the third
grade ELA test is adequate, with Cronbach’s alpha and the Stratified alpha both revealing a .93 reliability coefficient. The English Language Arts (ELA) and Math portions of the LEAP test are required for promotion from 4th to 5th grade. In order to pass to the next grade, students must achieve a combination of at least Approaching Basic level on one part of the exam and at least Basic on the other portion of the test (LEAP and GEE Interpretive Guide, Louisiana Department of Education, 2009).

**Student Level Variable**

**Previous Year Achievement**

Previous year academic achievement is used as a variable at the student level. Each student’s present score was subtracted from their prior score and a difference score was used.

**School Level Variables**

**Percent Free and Reduced Lunch**

Percent of school with free and reduced lunch is used as a variable at the school level. The percent of students receiving free and reduced lunch is used.

**Previous Year Achievement**

Previous year academic achievement is used as a variable at the school level. The average LEAP or iLEAP score for the 2007-2008 school year is used.

**PBIS Implementation Level**

The level of PBIS implementation is a variable used at the school level. Two different analyses were done using level of implementation. The first one used the self-assessment BOQ score. This was coded as 1 (low-implementing, score <70%) or 2 (high-implementing, score >= 70%). The second one used the external assessment SET score. This was coded as 1 (low-
implementing, score <80%) or 2 (high-implementing, score >= 80% overall and 80% on the B component; Horner et al., 2004).

**Suspensions and Expulsions**

**In School Suspensions (ISS).** The rate of students per 100 students where the disciplinary action is exclusion within the school building of a student from the student’s regular education program for up to but not more than 10 school days for disciplinary reasons by the school principal.

**Out of School Suspensions (OSS).** The rate of students per 100 students where the disciplinary action is exclusion outside of the school building of a student from the student’s regular education program for up to but not more than 10 school days for disciplinary reasons by the school principal.

**In School Expulsions (ISE).** The rate of students per 100 students where the disciplinary action is exclusion inside of the school building of a student from the student’s regular education program for more than 10 school days for disciplinary reasons by the school principal. There is usually a hearing examiner involved in expulsions.

**Out of School Expulsions (OSE).** The rate of students per 100 students where the disciplinary action is exclusion outside of the school building of a student from the student’s regular education program for more than 10 school days for disciplinary reasons by the school principal. There is usually a hearing examiner involved in expulsions.

**Analysis Strategy**

The analysis of data collected for this project can be divided into three phases: Data Quality, Model Assumptions, and The Model.
Data Quality

Using SPSS, each variable was inspected to ensure that the correct columns have been read and that cases with missing values have been deleted. The consistency was checked with data records, as well as the number of observations compared with the number in record. As the sample size is fairly large, all cases with missing data on the student level were deleted, as long as these cases were small number and random. The number of cases deleted are reported in the analysis. Each variable in the dataset was graphed and inspected for problematic data elements such as outliers, and all assumptions of the model were checked.

Model Assumptions

Assumptions of the Multilevel Model are:

1. **Linearity:** function forms are linear at each level. A scatterplot was generated and inspected for the general tendency of the data. A trend analysis was also run to check for linearity.

2. **Normality:** level-1 residuals are normally distributed and level-2 random effects u’s have a multivariate normal distribution. One of the major assumptions of the tests of significance used in the multilevel programs is normality of the error distributions involved. In a study by Maas and Hox (2004), the number of groups, the group size, and the intraclass correlation, with the second level residual errors following one of three non-normal distributions were varied. In addition asymptotic maximum likelihood standard errors are compared to robust (Huber/White) standard errors. The results from Maas and Hox (2004) show that non-normal residuals at the second level of the model have little or no effect on the parameter estimates. For the fixed parameters,
both the maximum likelihood-based standard errors and the robust standard errors are accurate. For the parameters in the random part of the model, the maximum likelihood-based standard errors at the lowest level are accurate, while the robust standard errors are often overcorrected. The standard errors of the variances of the level-two random effects are highly inaccurate, although the robust errors do perform better than the maximum likelihood errors. For good accuracy, robust standard errors need at least 100 groups. Thus, using robust standard errors as a diagnostic tool seems to be preferable to simply relying on them to solve the problem (Maas & Hox, 2004).

3. **Homoscedasticity: level-1 residual variance is constant.** This was tested on SPSS using scatterplots and box and whisker plots. Multi-level model assumes that scatter of points around the best-fit line has the same standard deviation all along the curve (homoscedasticity). The assumption is violated if the points with high or low X values tend to be further from the best-fit line.

4. **Independence: level-1 residuals and level-2 residuals are uncorrelated.** A correlation was run to see if the level1 and level 2 residuals are uncorrelated.

5. **Independence: observations at highest level are independent of each other.** To check for lack of independence, an ordinary least-squares (OLS) regression was run and saved the residuals. An ANOVA of residuals by group will also be run. If the ANOVA F-test is significant, the null hypothesis will be rejected that residuals are independent by group. That is, a significant F means data are correlated, not independent, and multi-level model should be used instead of OLS regression.
6. **Adequate sample size.** Hox (1995) suggests that for multi-level models, the higher level sample size be at least 20, preferably 50, and if variance components are important, preferably 100. For structural equation modeling approaches, Hox recommended sample size should be at least 50, preferably 100.

Should any of these assumptions not be met, then appropriate adjustments will be made.

**The Model**

To understand multi-level models, it is useful to begin by considering a simple (single-level) regression model. The formula for a simple regression model is given by:

\[ y_i = b_0 + b_1 x_i + e_i \]  

(1)

where subscript \( i \) takes values from 1 to \( n \), and where \( n \) is the total number of students in the sample of schools. For the \( i \)-th student, \( y_i \) is the outcome variable (the mean LEAP or iLEAP score) \( x_i \) is the control variable.

A student's predicted score \( (Y_i) \) is given by:

\[ Y_i = b_0 + b_1 x_i \]  

(2)

where \( Y_i \) is the predicted examination score for the \( i \)-th student, \( b_0 \) is the intercept where the regression line meets the vertical axis, and \( b_1 \) is the slope.

In equation (1), \( e_i \) is the difference between the \( i \)-th student's actual examination score and the predicted score. It is referred to as a 'residual' because it is that part of the score that is not predicted by the fixed part of the model represented by equation (2).
Moving now to a simple multi-level model, we can rewrite equation (1) as

\[ y_{ij} = b_{0j} + b_{1j}x_{ij} + e_{ij} \]  \hspace{1cm} (3)

in which terms are denoted by two subscripts, one indicating the student and the other the school attended by the student. It is likely that there will be variation among schools in the parameters of the model. For example, if \( b_{1j} \) is the slope which relates prior academic achievement to current academic achievement, \( y_{ij} \), then it can be expected that schools will vary in the strength of this relationship. This variability can be modeled as a function of school level characteristics

\[ b_{1j} = M + B^*W_j + u_j \]  \hspace{1cm} (4)

\( W \) in Equation 4 can represent teacher characteristics, characteristics of the student population at a school or whether or not a school participates in the PBIS program. Equations 3 and 4 constitute a multilevel model. It is the existence of the two residuals - the level-2 school residual and the level-1 student residual - which identifies this as a multi-level model.

The model described above helps capture the relationship between the covariates and the initial status, as well as, the covariates and growth rates. The variance estimates of the intercept and slope can be compared to the unconditional or to other nested models to see if the fitting of the covariates improved the fits. Interactions between previous year achievement and various level-2 factors, as well as interactions among certain level-2 factors can be studied.
Hypothesis testing of the level 2 model was done without out PBIS implementation level, and then add PBIS implementation level to see if there was an effect.

For the current analysis, an indicator variable was used in Equation 4 to represent participation status, and other school level covariates were treated as control variables. Equation 3 consisted of student level variables considered relevant to current academic achievement and all will be centered so that the slope in Equation 3 can be interpreted as the mean test results for a given school. The primary analysis focused on whether or not there was evidence that full implementation of universal PBIS helps predict student achievement once appropriate controls have been considered.
CHAPTER 4: DATA ANALYSIS

Preliminary data screening was done after the data were received from the Louisiana Department of Education. The data were encrypted as to protect it and also password protected although there were no identifying information in the data. Each student in the data files was given a SRAA, a random identifier so they could be matched across the years. Spring 2007, 2008, and 2009 test scores were obtained for both iLEAP and LEAP tests.

Upon first inspection of the data, there were approximately 215 variables and 250,000 entries for the iLEAP scores each year and 420 variables and 100,000 entries for the LEAP scores for each of the three years. The percentage of available SET scores from middle and high schools across Louisiana were very low, in comparison to elementary schools. Due to the majority of the SET scores being from elementary schools and in reference to the literature by Childs, Kincaid, and George (2010) that the SET and BOQ were more consistent across the elementary levels, only the scores for 3rd, 4th, and 5th graders were kept for the analysis. Also, many of the variables in the original data set were deleted as they will not be used in the analysis or were filler variables. All homestudy site codes were deleted from the analysis, as this study is focusing on public schools in Louisiana.

Upon first inspection of the data, there were 2,028 non-valid or missing scores from the LEAP scores. Those were excluded as it is a small percentage (3.5%) of the included cases. Some additional data was excluded from the analysis, but further explanation of those processes is included as each variable is excluded.
Data Quality

Using SPSS, each variable was inspected to ensure that the correct columns have been read and that cases with missing values have been deleted. Scores were verified for a few schools and districts using the school’s report cards and the data used in the analysis. For Spring 2008 LEAP scores, site code: 032003 was compared and showed for the SPSS data, which is shown in Table 1.

Table 1. Spring 2008 LEAP Scores for Site Code 032003 on Data Set

<table>
<thead>
<tr>
<th>Advanced</th>
<th>Approaching Basic</th>
<th>Basic</th>
<th>Mastery</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>41</td>
<td>28</td>
<td>7</td>
</tr>
</tbody>
</table>

The data set used for this analysis was also compared with the information on the school report card as illustrated in Table 2.

Table 2. Spring 2008 LEAP Scores for Site Code 032003 on School Report Card

<table>
<thead>
<tr>
<th>Advanced</th>
<th>Approaching Basic</th>
<th>Basic</th>
<th>Mastery</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>41</td>
<td>27</td>
<td>6</td>
</tr>
</tbody>
</table>

A sample of 5% of the cases were printed (equal numbers from the beginning, middle and end) and all data elements were inspected. For each variable in the dataset, the maximum and minimum scores were printed and compared to the range of possible values. A frequency distribution was generated for each variable in the dataset. These were inspected for unusual values or outliers. There were no unusual values or outliers.
**Data Comparison**

There are 54,896 students for Louisiana State DOE table for LEAP data. On the SPSS file there are 56,926 cases only using 4\textsuperscript{th} grade scores minus 2,028 non-valid or missing scores equal 54,898 valid scores in the chart below. These were compared with the Louisiana State Department of Education 2008 LEAP test scores in the chart below, using the achievement levels of Advanced (A), Mastery (M), Approaching Basic (AB), Basic (B), and Unsatisfactory (U).

According to the Louisiana state Department of Education published test scores matched with the SPSS files used in the data analysis, the data matched up within 13 cases out of over 54,000 cases, as illustrated in Table 3.

**Table 3. English Language Arts: Number of Students at Each Achievement Level**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>A</th>
<th>M</th>
<th>B</th>
<th>AB</th>
<th>U</th>
<th>Total Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Louisiana DOE File</td>
<td>2,089</td>
<td>11,721</td>
<td>24,286</td>
<td>10,298</td>
<td>6,502</td>
<td>54,896</td>
</tr>
<tr>
<td>SPSS File</td>
<td>2,088</td>
<td>11,720</td>
<td>24,281</td>
<td>10,294</td>
<td>6,500</td>
<td>54,883</td>
</tr>
</tbody>
</table>

Demographic analysis was run for the 3 years of test scores. The results are shown in Table 4.

**Research Question #1**

Is the degree of implementation of the universal level of PBIS related to changes in rates of discipline actions at the school level?

Correlations were done for preliminary school level runs to correlate high SET score with the BOQ score from 2007 and 2008. Correlations show that a school’s high SET score is correlated with a BOQ score in 2007 and 2008. Also 2007 BOQ scores correlate with 2008 BOQ
Table 4. Demographics for Data Set with All 3 Years of Test Scores

<table>
<thead>
<tr>
<th></th>
<th>2007 iLEAP 3rd Grade</th>
<th>2008 LEAP 4th grade</th>
<th>2009 iLEAP 5th grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>50,567</td>
<td>57,427</td>
<td>47,693</td>
</tr>
<tr>
<td>Matched Data</td>
<td>25,969</td>
<td>25,868</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>25,886 (51.2%)</td>
<td>29,783 (51.9%)</td>
<td>24,089 (50.5%)</td>
</tr>
<tr>
<td></td>
<td>13,253 (51%)</td>
<td>13,048 (50.4%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>24,502 (48.5%)</td>
<td>27,957 (48%)</td>
<td>23,517 (49.3%)</td>
</tr>
<tr>
<td></td>
<td>12,643 (48.7%)</td>
<td>12,804 (49.5%)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>179 (.4%)</td>
<td>65 (0.1%)</td>
<td>87 (.2%)</td>
</tr>
<tr>
<td></td>
<td>73 (0.3%)</td>
<td>16 (0.1%)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>24,966 (49.4%)</td>
<td>26,907 (46.9%)</td>
<td>23,968 (50.3%)</td>
</tr>
<tr>
<td></td>
<td>13,373 (51.5%)</td>
<td>13,533 (52.3%)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>23,173 (45.8%)</td>
<td>27,803 (48.4%)</td>
<td>21,178 (44.4%)</td>
</tr>
<tr>
<td></td>
<td>11,196 (43.1%)</td>
<td>10,895 (42.1%)</td>
<td></td>
</tr>
<tr>
<td>Free and Reduced Lunch Status</td>
<td>34,525 (68.3%)</td>
<td>40,357 (70.2%)</td>
<td>31,690 (66.4%)</td>
</tr>
<tr>
<td></td>
<td>17,183 (66.2%)</td>
<td>16,972 (65.6%)</td>
<td></td>
</tr>
<tr>
<td>Full Price</td>
<td>15,856 (31.4%)</td>
<td>16,679 (29%)</td>
<td>15,864 (33.3%)</td>
</tr>
<tr>
<td></td>
<td>8,727 (33.6%)</td>
<td>8,873 (34.3%)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>186 (.4%)</td>
<td>391 (0.7%)</td>
<td>139 (.3%)</td>
</tr>
<tr>
<td></td>
<td>59 (0.2%)</td>
<td>23 (0.1%)</td>
<td></td>
</tr>
</tbody>
</table>

scores, as illustrated in Table 5. This finding was also replicated in ULL report with scores from the 2007-2008 school year using SET score and BOQ correlation \( r=0.547; N = 92 \) Schools; Mean SET = 86.8; Mean BOQ = 73.4).

The average suspension & expulsion rates were correlated with all PBIS implementation scores. In 2007, the only statistically significant result was in the category of out of school
suspensions correlated with 2007 BOQ scores and high SET score. In 2008, there are statistically significant correlations between high SET score and out of school suspension rate and in school expulsion rate. There are no statistically significant correlations between BOQ scores in 2008 and disciplinary actions as in the previous year.

A score was created that represents the change in average suspension & expulsion rates from 2007 to 2008. There were no correlations between this change score and any of the PBIS implementation measures.

A score was created that represents the average z-score change in achievement in all content areas and it was correlated to all PBIS implementation scores. In reading, the average z-score change in achievement was correlated with the high SET score and 2007 BOQ score. In math, science and social studies, this score showed a statistically significant correlation with only the high SET score. There were no correlations with the ELA score.

The change in discipline rates in all 4 categories from 2007 to 2008 were correlated with the average z-score change in achievement in all 5 content areas. There are statistically significant correlations between change in out of school suspension rates with all 5 content area scores, and the change in out of school expulsion rates with all content areas except for science. All of these correlations are shown in Table 6.

### Table 5. Pearson Correlation for High SET Score, 2007 BOQ, and 2008 BOQ

<table>
<thead>
<tr>
<th></th>
<th>2007 BOQ score</th>
<th>2008 BOQ score</th>
</tr>
</thead>
<tbody>
<tr>
<td>High SET score</td>
<td>.414** (n=371)</td>
<td>.171** (n=456)</td>
</tr>
<tr>
<td>2007 BOQ score</td>
<td>1**</td>
<td>.510** (n=431)</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level
Table 6. Summary of Correlations

<table>
<thead>
<tr>
<th></th>
<th>HighSET</th>
<th>BOQ 2007</th>
<th>BOQ 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HighSET</strong></td>
<td>R = 1</td>
<td>R = .414**</td>
<td>R = .171**</td>
</tr>
<tr>
<td></td>
<td>N = 728</td>
<td>N = 371</td>
<td>N = 456</td>
</tr>
<tr>
<td><strong>BOQ 2007</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2007</td>
<td>2008</td>
<td>R = .510**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 431</td>
</tr>
<tr>
<td><strong>ISS Q7/08</strong></td>
<td>R = .010</td>
<td>R = -.050</td>
<td>R = -.075</td>
</tr>
<tr>
<td></td>
<td>N = 584</td>
<td>N = 596</td>
<td>N = 387</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 511</td>
</tr>
<tr>
<td><strong>OSS 07/08</strong></td>
<td>R = -.196**</td>
<td>R = -.133**</td>
<td>R = -.079</td>
</tr>
<tr>
<td></td>
<td>N = 657</td>
<td>N = 652</td>
<td>N = 476</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 558</td>
</tr>
<tr>
<td><strong>ISE 07/08</strong></td>
<td>R = -.117*</td>
<td>R = -.168**</td>
<td>R = -.075</td>
</tr>
<tr>
<td></td>
<td>N = 312</td>
<td>N = 380</td>
<td>N = 226</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 345</td>
</tr>
<tr>
<td><strong>OSE 07/08</strong></td>
<td>R = -.046</td>
<td>R = .071</td>
<td>R = -.039</td>
</tr>
<tr>
<td></td>
<td>N = 220</td>
<td>N = 183</td>
<td>N = 180</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N = 161</td>
</tr>
<tr>
<td><strong>CHANGE ISS</strong></td>
<td>R = -.045</td>
<td>R = -.039</td>
<td>R = -.056</td>
</tr>
<tr>
<td></td>
<td>N = 542</td>
<td>N = 353</td>
<td>N = 460</td>
</tr>
<tr>
<td><strong>CHANGE OSS</strong></td>
<td>R = .058</td>
<td>R = -.029</td>
<td>R = -.038</td>
</tr>
<tr>
<td></td>
<td>N = 634</td>
<td>N = 452</td>
<td>N = 543</td>
</tr>
<tr>
<td><strong>CHANGE ISE</strong></td>
<td>R = -.027</td>
<td>R = -.094</td>
<td>R = .019</td>
</tr>
<tr>
<td></td>
<td>N = 251</td>
<td>N = 179</td>
<td>N = 252</td>
</tr>
<tr>
<td><strong>CHANGE OSE</strong></td>
<td>R = -.062</td>
<td>R = .027</td>
<td>R = -.032</td>
</tr>
<tr>
<td></td>
<td>N = 129</td>
<td>N = 107</td>
<td>N = 107</td>
</tr>
<tr>
<td><strong>GRAND MEAN READ</strong></td>
<td>R = .127*</td>
<td>R = .140*</td>
<td>R = .037</td>
</tr>
<tr>
<td></td>
<td>N = 384</td>
<td>N = 212</td>
<td>N = 260</td>
</tr>
<tr>
<td><strong>GRANDMEAN MATH</strong></td>
<td>R = .122*</td>
<td>R = .104</td>
<td>R = -.010</td>
</tr>
<tr>
<td></td>
<td>N = 384</td>
<td>N = 212</td>
<td>N = 260</td>
</tr>
<tr>
<td><strong>GRANDMEAN ELA</strong></td>
<td>R = .093</td>
<td>R = .072</td>
<td>R = .052</td>
</tr>
<tr>
<td></td>
<td>N = 384</td>
<td>N = 212</td>
<td>N = 260</td>
</tr>
<tr>
<td><strong>GRANDMEAN SCI</strong></td>
<td>R = .131*</td>
<td>R = .100</td>
<td>R = -.034</td>
</tr>
<tr>
<td></td>
<td>N = 383</td>
<td>N = 212</td>
<td>N = 259</td>
</tr>
<tr>
<td><strong>GRANDMEAN SS</strong></td>
<td>R = .123*</td>
<td>R = .094</td>
<td>R = .036</td>
</tr>
<tr>
<td></td>
<td>N = 383</td>
<td>N = 212</td>
<td>N = 259</td>
</tr>
</tbody>
</table>

*SIGNIFICANCE at 0.05
**SIGNIFICANCE at 0.01
Research Question #2

Are measures of implementation (BOQ, SET) of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP?

Assumptions of the multilevel model were checked as described below: Checking the Assumptions of Multilevel Model:

1. **Linearity**: function forms are linear at each level. A scatterplot was generated and inspected for the general tendency of the data. A trend analysis was also run to check for linearity. The general tendency of the data was linear.

2. **Normality**: level-1 residuals are normally distributed and level-2 random effects u’s have a multivariate normal distribution. One of the major assumptions of the tests of significance used in the multilevel programs is normality of the error distributions involved. In a study by Maas and Hox (2004), the number of groups, the group size, and the intraclass correlation, with the second level residual errors following one of three non-normal distributions were varied. In addition asymptotic maximum likelihood standard errors are compared to robust (Huber/White) standard errors. The results from Maas and Hox show that non-normal residuals at the second level of the model have little or no effect on the parameter estimates. For the fixed parameters, both the maximum likelihood-based standard errors and the robust standard errors are accurate. For the parameters in the random part of the model, the maximum likelihood-based standard errors at the lowest level are accurate, while the robust standard errors are often overcorrected. The standard errors of the variances of the level-two random effects are highly inaccurate, although the
robust errors do perform better than the maximum likelihood errors. For good accuracy, robust standard errors need at least 100 groups. Thus, using robust standard errors as a diagnostic tool seems to be preferable to simply relying on them to solve the problem (Maas & Hox, 2004). In the current study, there are about 365 level 2 groups or schools.

3. **Homoscedasticity: level-1 residual variance is constant.** This was tested on SPSS using scatterplots and box and whisker plots. Multi-level model assumes that scatter of points around the best-fit line has the same standard deviation all along the curve (homoscedasticity). The points with high or low X values tend to be close to the best-fit line.

4. **Independence: level-1 residuals and level-2 residuals are uncorrelated.** A correlation was run to see if the level 1 and level 2 residuals are uncorrelated. The residuals were uncorrelated.

5. **Independence: observations at highest level are independent of each other.** To check for lack of independence, an ordinary least-squares (OLS) regression was run and the residuals were saved. An ANOVA of residuals by group was also run. The ANOVA F-test is significant, the null hypothesis was rejected that residuals are independent by group. A significant F means data are correlated, not independent, and multi-level model should be used instead of OLS regression.

6. **Adequate sample size: Hox (1995) suggests that for multi-level models, the higher level sample size be at least 20, preferably 50, and if variance components are**
important, preferably 100. The sample size for this study is 25,969 for Level 1 and 385 groups for Level 2.

The Model

Data was matched across student scores. The first match was done with 2007 iLEAP scores for grade 3 to 2008 LEAP scores from grade 4. When this match was done, there were 46,312 subjects matched across the 2 years. The descriptive statistics and correlations were run for this dataset, which are shown in Table 7.

**Table 7. Correlations and Descriptive Statistics for Grade 3 to Grade 4 Match**

<table>
<thead>
<tr>
<th></th>
<th>MathSS-G3</th>
<th>MathSS-G4</th>
<th>ReadingSS-G3</th>
<th>ReadingSS-G4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MathSS-G3</strong></td>
<td>1</td>
<td>.763</td>
<td>.679</td>
<td>.634</td>
</tr>
<tr>
<td><strong>MathSS-G4</strong></td>
<td>.763</td>
<td>1</td>
<td>.595</td>
<td>.699</td>
</tr>
<tr>
<td><strong>ReadingSS-G3</strong></td>
<td>.679</td>
<td>.595</td>
<td>1</td>
<td>.649</td>
</tr>
<tr>
<td><strong>ReadingSS-G4</strong></td>
<td>.634</td>
<td>.699</td>
<td>.649</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>305.14</td>
<td>59.526</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>SD</td>
<td>338.19</td>
<td>54.357</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>Minimum</td>
<td>301.67</td>
<td>68.159</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>Maximum</td>
<td>330.73</td>
<td>53.72</td>
<td>100</td>
<td>500</td>
</tr>
</tbody>
</table>

n=46,312

The PBIS school indicator score used in this study is the SET score, a school level score given to each school by an external evaluator. The original dataset of SET scores across the state of Louisiana included a total of 1,231 SET scores from 743 schools. The highest SET score from a school was taken as the SET score for the school. The highest SET score from each school was matched with the student data file for each school code. All subjects not associated
with a school that had a SET score were not included for this analysis. This brought the grade 3 to grade 4 matched data file from 46,312 students down to 25,969 students. The grade 4 to grade 5 matched dataset was reduced from 44,538 students to 25,868 students. Correlations and descriptive statistics for the reduced data set are shown in Table 8.

Table 8. Correlations and Descriptive Statistics for the Grade 3 to Grade 4 Match Only Including Students with SET School Scores

<table>
<thead>
<tr>
<th></th>
<th>MathSS-G3</th>
<th>MathSS-G4</th>
<th>ReadingSS-G3</th>
<th>ReadingSS-G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MathSS-G3</td>
<td>1</td>
<td>.766</td>
<td>.679</td>
<td>.637</td>
</tr>
<tr>
<td>MathSS-G4</td>
<td>.766</td>
<td>1</td>
<td>.599</td>
<td>.700</td>
</tr>
<tr>
<td>ReadingSS-G3</td>
<td>.679</td>
<td>.599</td>
<td>1</td>
<td>.656</td>
</tr>
<tr>
<td>ReadingSS-G4</td>
<td>.637</td>
<td>.700</td>
<td>.656</td>
<td>1</td>
</tr>
</tbody>
</table>

Mean                  306.72  339.14  302.89  330.71
SD                    59.532  54.456  68.215  53.512
Minimum               100     100      100      100
Maximum               500     500      500      500

n=25,969

The second match was done with 2008 LEAP scores for 4th grade to 2009 iLEAP scores for 5th grade. When this match was completed there were 44,538 subjects across the 2 years. The descriptive statistics and correlations were also run on this data set with the results below in Table 9 and for only schools with SET scores in Table 10.
Table 9. Correlations and Descriptive Statistics for Grade 4 to Grade 5 Match

<table>
<thead>
<tr>
<th></th>
<th>MathSS-G4</th>
<th>MathSS-G5</th>
<th>ReadingSS-G4</th>
<th>ReadingSS-G5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MathSS-G4</td>
<td>1</td>
<td>.757</td>
<td>.636</td>
<td>.553</td>
</tr>
<tr>
<td>MathSS-G5</td>
<td>.757</td>
<td>1</td>
<td>.571</td>
<td>.619</td>
</tr>
<tr>
<td>ReadingSS-G4</td>
<td>.636</td>
<td>.571</td>
<td>1</td>
<td>.626</td>
</tr>
<tr>
<td>ReadingSS-G5</td>
<td>.553</td>
<td>.619</td>
<td>.626</td>
<td>1</td>
</tr>
</tbody>
</table>

Mean                  | 345.86    | 307.00    | 337.66       | 302.35       |
SD                     | 46.438    | 61.849    | 45.912       | 60.749       |
Minimum                | 100       | 100       | 100          | 100          |
Maximum                | 500       | 500       | 500          | 500          |

n=44,538

Table 10. Correlations and Descriptive Statistics for the Grade 4 to Grade 5 Match Only Including Students with SET School Scores

<table>
<thead>
<tr>
<th></th>
<th>MathSS-G4</th>
<th>MathSS-G5</th>
<th>ReadingSS-G4</th>
<th>ReadingSS-G5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MathSS-G4</td>
<td>1</td>
<td>.759</td>
<td>.638</td>
<td>.558</td>
</tr>
<tr>
<td>MathSS-G5</td>
<td>.759</td>
<td>1</td>
<td>.573</td>
<td>.621</td>
</tr>
<tr>
<td>ReadingSS-G4</td>
<td>.638</td>
<td>.573</td>
<td>1</td>
<td>.632</td>
</tr>
<tr>
<td>ReadingSS-G5</td>
<td>.558</td>
<td>.621</td>
<td>.632</td>
<td>1</td>
</tr>
</tbody>
</table>

Mean                  | 346.54    | 308.44    | 337.32       | 303.46       |
SD                     | 46.881    | 62.064    | 46.124       | 60.642       |
Minimum                | 100       | 100       | 100          | 100          |
Maximum                | 500       | 500       | 500          | 500          |

n=25,868

Research Question #3

3. Do identified student/ school characteristics contribute to any difference in academic performance?
Multilevel Model

To see if there is enough variation between schools to warrant further analysis, an unconditional model was run for reading and math in both data sets, as shown in Table 11. From these analyses the intraclass correlation was computed. Small values of the ICC indicate that multilevel modeling is not necessary; however that is not the case for these 4 unconditional models. In each of these models there is about a 15% to 20% variance between schools. Below are the intraclass correlations for each of the 2 year analysis in both reading and math.

Table 11. Unconditional Model for Grade 3 to Grade 4 and Grade 4 to Grade 5 Match

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model G3 to G4 Reading</th>
<th>Unconditional Model G4 to G5 Reading</th>
<th>Unconditional Model G3 to G4 Math</th>
<th>Unconditional Model G4 to G5 Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted mean</td>
<td>416.2</td>
<td>567.99</td>
<td>590.76</td>
<td>733.17</td>
</tr>
<tr>
<td>Residual</td>
<td>2461.5</td>
<td>3135.07</td>
<td>2399.77</td>
<td>3149.549</td>
</tr>
<tr>
<td>ICC</td>
<td>0.1446</td>
<td>0.1533</td>
<td>0.1975</td>
<td>0.188</td>
</tr>
</tbody>
</table>

n=25.969

To fit the Grand Mean centered model, the approach taken was to adjust grade 4 means for each school for the performance of their students as 3rd graders. The same approach was taken to adjust the grade 5 means for each school for the performance as 4th graders. Grand mean centering was done with the reading and math scaled scores for each dataset containing the students matched across the 2 years containing their school SET scores. In the first dataset from 2007 to 2008 matched 3rd and 4th graders the grand mean centering took into account each student’s 3rd grade reading scaled score from which the 2007 mean for all 3rd graders
across the whole data set was then subtracted. This was also done for math, ELA, science, and social studies scaled scores. The additional data set from 2008-2009 also had scores computed for each student for grand mean centering all 5 scaled scores.

The multilevel model was then run for all of the subscores in the 5 subject areas with Model I using each student’s previous scale score as a covariate, Model II adding in percent free and reduced lunch at the school level as a covariate, and Model III adding the school SET score. This data and results are shown in Table 12 through Table 16 for the Grade 3 to grade 4 matched data set for reading (Table 12), math (Table 13), English language arts (Table 14), science (Table 15), and social studies (Table 16). This focuses on research questions 2 and 3:

2. Are measures of implementation (BOQ, SET) of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP?

3. Do identified student/school characteristics contribute to any difference in academic performance from 3rd grade iLEAP to 4th grade LEAP scores?

<table>
<thead>
<tr>
<th>Table 12. Reading: Grade 3 to Grade 4 Using High SET Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unconditional Model</strong></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Previous Scale Score</td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
</tr>
<tr>
<td>SET Score</td>
</tr>
</tbody>
</table>
Table 12. (Continued) Reading: Grade 3 to Grade 4 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance of the Intercept</td>
<td>416.19</td>
<td>81.35</td>
<td>62.59</td>
<td>61.27</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td>80.4%</td>
<td>23.0%</td>
<td>2.1%</td>
<td></td>
</tr>
</tbody>
</table>

n=25,969

Table 13. Math: Grade 3 to Grade 4 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>336.08 (1.29)*</td>
<td>338.05 (0.63)*</td>
<td>346.27 (2.03)*</td>
<td>341.38 (6.15)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.70 (0.004)*</td>
<td>0.70 (0.004)*</td>
<td>0.70 (0.004)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td>-0.12(0.02)*</td>
<td>-0.12 (0.03)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td>0.05 (0.06)</td>
<td></td>
</tr>
</tbody>
</table>

|                           |                     |             |
| Variance of the Intercept | 590.76              | 126.51      | 119.70     | 119.79     |
| % Change in Variance of the Intercept | 78.6%               | 5.3%        |

n=25,969
* Significant at the 0.01 level
### Table 14. ELA: Grade 3 to Grade 4 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>322.05 (1.16)*</td>
<td>323.56 (0.55)*</td>
<td>332.42 (1.77)*</td>
<td>321.96 (5.37)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.69 (0.004)*</td>
<td>0.69 (0.004)*</td>
<td>0.69 (0.004)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.12 (0.02)*</td>
<td>-0.12 (0.02)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.11 (0.05)*</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>465.71</td>
<td>93.69</td>
<td>85.48</td>
<td>84.39</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td>79.9%</td>
<td>8.8%</td>
<td>1.3%</td>
<td></td>
</tr>
</tbody>
</table>

n=25,969

* Significant at the 0.01 level

### Table 15. Science: Grade 3 to Grade 4 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>316.03(1.29)*</td>
<td>318.02(0.63)*</td>
<td>340.22(1.72)*</td>
<td>327.16(5.20)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.66 (0.004)*</td>
<td>.65 (0.004)*</td>
<td>0.65 (0.004)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.32 (0.02)*</td>
<td>-0.31 (0.23)*</td>
<td>0.14 (0.51)*</td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>600.79</td>
<td>128.54</td>
<td>79.72</td>
<td>78.56</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td>78.6%</td>
<td>37.9%</td>
<td>1.5%</td>
<td></td>
</tr>
</tbody>
</table>

n=25,969

* Significant at the 0.01 level
Table 16. Social Studies: Grade 3 to Grade 4 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>312.69 (1.09)*</td>
<td>314.19 (0.59)*</td>
<td>335.44 (1.60)*</td>
<td>321.40 (4.85)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>.55 (0.004)*</td>
<td>0.55 (0.004)*</td>
<td>0.55 (0.004)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.30 (0.02)*</td>
<td>-0.29 (0.02)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.15 (0.05)*</td>
</tr>
<tr>
<td>Variance of the</td>
<td>426.17</td>
<td>118.05</td>
<td>71.67</td>
<td>70.02</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Change in</td>
<td>72.3%</td>
<td>39.3%</td>
<td>2.3%</td>
<td></td>
</tr>
<tr>
<td>Variance of the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n=25,969
* Significant at the 0.01 level

All of the models showed significance at the 0.01 level for a school’s SET score except for math. In Reading, ELA, Science, and Social Studies the SET score accounted for about 1 to 2% of the change in variance of the intercept.

The Model for the Grade 4 to Grade 5 Matched Data SET

The same model was run for the grade 4 to grade 5 student matched test scores. Table 17 through Table 21 show the results for the grade 4 to grade 5 matched data set for reading (Table 17), math (Table 18), English language arts (Table 19), science (Table 20), and social studies (Table 21). This focuses on research questions 2 and 3.
2. Do identified student/school characteristics contribute to any difference in academic performance from 4th grade LEAP to 5th grade iLEAP scores?

3. Is a full level of implementation of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP from grade 4 to grade 5?

Table 17. Reading: Grade 4 to Grade 5 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>299.13 (1.27)*</td>
<td>301.86 (0.76)**</td>
<td>332.52 (1.84)**</td>
<td>327.18 (5.79)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.78 (0.01)**</td>
<td>0.767(0.01)**</td>
<td>0.77(0.01)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td>-0.43 (0.03)**</td>
<td>-0.43(0.03)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.05(0.06)</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>567.99</td>
<td>185.22</td>
<td>84.08</td>
<td>84.24</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td>67.4%</td>
<td>54.6 %</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n= 25,868  
* Significant at the 0.01 level
Table 18. Math: Grade 4 to Grade 5 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>303.28 (1.43)*</td>
<td>307.28 (0.75)*</td>
<td>333.02 (2.07)*</td>
<td>331.76 (6.26)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.97 (0.01)*</td>
<td>0.96 (0.01)*</td>
<td>0.96 (0.01)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.37 (0.03)*</td>
<td>-0.37 (0.03)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.01 (0.06)</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>733.17</td>
<td>191.86</td>
<td>122.54</td>
<td>122.89</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td></td>
<td>73.8%</td>
<td>36.1%</td>
<td></td>
</tr>
</tbody>
</table>

n= 25,868
* Significant at the 0.01 level

Table 19. ELA: Grade 4 to Grade 5 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>298.96 (1.09)*</td>
<td>301.95 (0.54)*</td>
<td>319.66 (1.52)*</td>
<td>318.93 (4.69)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.77 (0.01)*</td>
<td>0.77 (0.01)*</td>
<td>0.77 (0.01)*</td>
<td></td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.25 (0.02)*</td>
<td>-0.25 (0.02)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.01 (0.04)</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>419.41</td>
<td>96.05</td>
<td>63.52</td>
<td>63.75</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td></td>
<td>77.1%</td>
<td>33.9%</td>
<td></td>
</tr>
</tbody>
</table>

n= 25,868
* Significant at the 0.01 level
### Table 20. Science: Grade 4 to Grade 5 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>299.21 (1.22)*</td>
<td>302.60 (0.59)*</td>
<td>323.42 (1.63)*</td>
<td>322.36 (4.99)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.74 (0.01)*</td>
<td>0.73 (0.01)*</td>
<td>0.73 (0.01)*</td>
<td>0.73 (0.01)*</td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.29 (0.02)*</td>
<td>-0.29 (0.02)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.01 (0.05)</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>540.58</td>
<td>115.40</td>
<td>72.23</td>
<td>72.47</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td></td>
<td>78.7%</td>
<td>37.4%</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>25,868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Significant at the 0.01 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 21. Social Studies: Grade 4 to Grade 5 Using High SET Score

<table>
<thead>
<tr>
<th></th>
<th>Unconditional Model</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>297.10 (1.01)*</td>
<td>299.66 (0.55)*</td>
<td>315.41 (1.61)*</td>
<td>308.42 (4.91)*</td>
</tr>
<tr>
<td>Previous Scale Score</td>
<td>0.68 (0.01)*</td>
<td>0.67 (0.01)*</td>
<td>0.67 (0.01)*</td>
<td>0.67 (0.01)*</td>
</tr>
<tr>
<td>%Free/Reduced Lunch</td>
<td></td>
<td>-0.22 (0.02)*</td>
<td>-0.22 (0.02)*</td>
<td></td>
</tr>
<tr>
<td>SET Score</td>
<td></td>
<td></td>
<td></td>
<td>0.07 (0.05)</td>
</tr>
<tr>
<td>Variance of the Intercept</td>
<td>364.72</td>
<td>97.19</td>
<td>72.26</td>
<td>72.02</td>
</tr>
<tr>
<td>% Change in Variance of the Intercept</td>
<td></td>
<td>73.4%</td>
<td>25.7%</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>25,868</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Significant at the 0.01 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the models run on the 4th to 5th grade matched data, there were no statistically significant results when the high SET score was added to the model.
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

The overall purpose of this study was to examine if implementation of the universal level of PBIS is related to student achievement on the LEAP and iLEAP examinations administered as part of the Louisiana Educational Assessment Program. A second purpose was to examine whether identified student/school characteristics contribute to any difference in academic performance. Specifically, the study addressed the following objectives:

1) Is the degree of implementation of the universal level of PBIS related to changes in rates of discipline actions at the school level?

2) Is a full level of implementation of the universal level of PBIS related to students’ achievement on the LEAP and iLEAP?

3) Do identified student/school characteristics contribute to any difference in academic performance?

The current research targeted students across the state of Louisiana in grades 3-5 for the years 2007-2009 in schools that had a SET score, an external assessment which measures PBIS implementation. However, the accessible population was students in 3rd grade in the spring of 2007 and continued through 4th grade in the spring of 2008, and students in 4th grade in spring of 2008 and continued through 5th grade in the spring of 2009; in addition the students had to be from schools that had a SET score from these years. This consisted of 25,969 students across Louisiana from 2007 to 2008 and 25,868 students across Louisiana from 2008 to 2009. This was reduced from the original approximately 55,000 students from each year in the original data set.
In terms of discipline rates in schools and implementation measures of PBIS, there were statistically significant correlations in 2007 between high SET score and out of school suspension rates paired with statistically significant correlations between BOQ scores and out of school suspension rates are consistent with other findings. Simonsen et al. (2012) state that, typical school-level disciplinary data were collected (i.e., ODR, OSS, TS), which often required an administrator’s response or action for data to be recorded (ODR) or generated (OSS and TS). Simonsen et al. (2012) states, “In other words, an administrator decides whether to enter an ODR into the data base and issue a suspension; therefore, these metrics may be underestimates of the actual levels of problem behavior in a school. To provide a greater depth of understanding of the specific effects of SWPBS implementation, future research should collect and examine information related to more immediate behavioral and social climate variables (e.g., academic engagement, opportunities to respond, scheduled/allocated academic time, and disruptive behavior) (p. 13).”

Simonsen and colleagues call for additional research and the challenge of securing data that consistently reflect the outcomes and processes of implementation are supported in the current study. Large scale implementation research can be complicated by factors beyond the control of the researchers and the program implementers.

Starting in 2008 schools were mandated by Louisiana Department of Education to submit BOQ scores to the state DOE. The number of schools submitting BOQ scores increased from 524 schools in Spring of 2007 to 658 in Spring of 2008, and finally to 1012 schools in Spring of 2009. When schools are mandated to submit scores to the state for accountability purposes the number of schools submitting increased by about 50% and the average score increased also. This may play some factor in the finding using the 2008 self-assessment scores.
In terms of academic outcomes, there are statistically significant correlations between change in out of school suspension rates with all 5 content area scores, and the change in out of school expulsion rates with all content areas except for science.

In summary, high SET scores have statistically significant correlations with out of school suspension rates in 2007 and 2008, in school expulsion rates in 2007 and 2008, and the average z-score change in academic achievement in reading, math, science, and social studies. BOQ 2007 scores had statistically significant correlations with only out of school suspension rates in 2007 and the average z-score change in academic achievement in reading. BOQ scores in 2008 had no statistically significant correlations between any of the outcome measures. Simonsen et al. (2012) found that regardless of fidelity, schools experienced decreases in ODRs over time, but time alone did not lead to overall decreases in OSS or total suspensions. The authors stated that one explanation may be that student behaviors that resulted in an ODR may be sensitive to change and responsive to SWPBS implementation, regardless of fidelity.

Using the consistent correlations between high SET scores for a school and various outcome measures, the multilevel model was run using high SET score as a factor for each of the schools. When controlling for a school’s percent of free and reduced lunch and each student’s prior test scores, there was a statistically significant result when adding in a school’s high SET score for all of the subject areas, except for math from 2007 to 2008. All of the models showed significance at the 0.01 level for a school’s SET score except for math. In Reading, ELA, science, and social studies the SET score accounted for about 1 to 2% of the change in variance of the intercept. For the models run on the 4th to 5th grade matched data, there were no statistically significant results when the high SET score was added to the model. Findings from
Simonsen, et al. (2012) agree that fidelity of implementation does improve academic outcome for students in math; whereas the findings of this study show that fidelity of implementation accounted for improvement in all subject areas except for math in grades 3-4, but not in grades 4-5. The model that was run in this study, however, was very different than the one run by Simonsen, et al. (2012) in reference to the levels for the multilevel model. The Illinois model was schools nested within districts, whereas this current study was students nested within schools, which may account for the differences in findings.

Conclusions, Implications, and Recommendations

One objective of this study was to determine the relationship between discipline rates of PBIS implementation measures. Overall, a high level of PBIS implementation was correlated with out of school suspension rates and in school expulsion rates and a school’s change in academic scores in reading, math, science and social studies. The research in this area is consistent with previous findings and the logic behind PBIS. Consistent with Simonsen et al. (2012) in reference to the recent finding with Illinois schools, SWPBS primarily supports a structure for improvement in social behavior; therefore, it was not surprising to find that schools that implemented SWPBS with fidelity also experienced lower rates of ODRs, OSS, and total suspensions.

The present study sought to determine if a high level of implementation of PBIS at the school level had an impact on a student’s academic scores based on high stakes test scores. Across the state of Louisiana, a schools’ implementation level of PBIS did show evidence of change of a student’s high-stakes test score from 2007 to 2008. However, it did not show a change from 2008 to 2009. In both 2 year spans, a previous year’s academic test score and a
school’s percent of free and reduced lunch were both predictors in the model. In 2007-2008, a school’s SET score was also a significant predictor in the model in all of the content areas, except for math. The findings from the multilevel model are consistent with other findings for the 3rd grade to 4th grade match. However, for the 4th to 5th grade match the absence of similar findings may reflect the fact that implementation scores were mandated at the state level for these years, which may have caused inflation in scores from each school. Even in the correlations, the 2008 scores were not statistically significant as the 2007 scores were found to be.

As a statewide study in Louisiana of the relationship between implementation level of PBIS and a student-level of academic achievement, the findings appear promising. In Louisiana, over 1,000 schools have been trained in PBIS and many of the schools have a high level of implementation. Horner et al., (2005) point out that academic and behavior supports must be intertwined. The importance of effective direct instruction in academic skills is critical to improving academic skills. Students will not learn academic skills without effective instruction and a good curriculum. This “good” curriculum is often tied to RTI research. Sugai, Horner, Fixsen, and Blase (2010) summarized that the PBIS framework is just the application of RTI principles to the improvement of social behavior outcomes for all students. PBIS is often described as the “behavior side” of the RTI multi-tiered continuum or triangle; however, this description misrepresents the actual integrated implementation of behavior and academic supports (Sugai et al., 2010). They will not learn to read just being taught social skills. Of course, these same students will not learn to read in a school or classroom that is behaviorally chaotic. In order to have students receive an effective education we need effective behavior
support interventions, an empirically validated curriculum as well as effective instruction. As Horner, Sugai, and Anderson (2010) state,

It is premature to claim that investing in SWPBS is causally associated with improved academic outcomes. In fact, the conceptual logic does not support the expectation that building social support would lead to improved reading, math, or writing skills. Rather, the expectation is that establishing a predictable, consistent, positive, and safe social culture will improve the behavioral engagement of students in learning, and that if this engagement is coupled with a functional curriculum and effective teaching, academic outcomes will be more likely (p. 8).

This study may well lend support to the assertion that the focus on a systematic approach to developing predictable and safe school environments (e.g. PBIS) may have fortuitous side effects. That is, a consistent, although small magnitude change in certain academic outcomes. By having schools adopt more consistent practices and providing feedback (e.g. SETs, BOQs) small but consistent changes may be expected in the absence of direct intervention. This supports Mayer’s (1995) contention of the importance of context in behavioral intervention.

The limitations of this study are issues with subject attrition and selective sampling. In reference to subject attrition, students had to be matched across 2 consecutive years at the same school. So in this study, students who repeated a grade or changed schools within the 2 year time period were not included in the data. This excluded about 40% of the original data sample, and possibly caused the data to not be as normally distributed without these scores. Many of the students who are transient are the same students that need more behavioral and academic support. Another issue relates to selective sampling. The SET scores used as the implementation measure of PBIS were also high (>80%) for the majority of the schools. Recommendations are that future replication studies be conducted with additional years of
testing data and PBIS implementation scores. These PBIS implementation scores are so important as they are the measure of treatment integrity. Sanetti, Dobey, and Gritter (2012) state that researchers must ensure that interventions are replicated and validated across different settings and populations, and that their essential components are determined. Advancing the field of PBIS, then, requires a focus on treatment integrity data. Both researchers and practitioners can use the tools available to them to contribute to this effort by developing treatment integrity assessments, keeping data on treatment integrity, and reporting these data as appropriate.

As in many large-scale studies using extant data sets this study was hampered by pragmatic concerns of various state and district administration. Certainly the inability/unwillingness to require individual schools to report discipline referral data restricts the sensitivity of measures available. The policy decision to mandate the reporting of BOQs for all districts as opposed to those voluntarily participating may have had unintended side effects (e.g. overinflated self-reports) that could limit the generalizability of these findings.

For future research, this study created an initial way to analyze large amounts of student-level data across a state combined with fidelity of implementation of PBIS. The findings are promising to show a link between a high level of PBIS implementation and behavioral and academic outcomes at the student level. Additional research should continue to employ multilevel modeling as one of several available statistical analyses. Multilevel modeling, as employed, in the present study allows a more direct analysis of how individual student level data can offer more comprehensive understanding of program or intervention outcomes. Another focus of research should be building level implementation and look at the factors and
barriers that attribute to this full implementation. Implementation is such an important key in the effect of any intervention at the school level, and the point at which full implementation occurs. In future replication studies, including ODRs and suspensions and expulsions, particularly at the student level may also show additional findings.

Continued research should focus on implementation as a key area of research to develop more efficient and effective protocols for supporting and measuring implementation, to identify additional variables that promote or hinder implementation, and to identify the threshold of implementation necessary to produce program outcomes. Research should continue to address the interaction between social and academic behavior (Kaufman, 2011). Understanding not only the extent of reciprocity between the two behavioral domains but also how these may be affected by various levels and focus of interventions remain areas for fruitful research.

Additional research should also address the development new or improvement of existing measures of implementation. These procedures for measuring implementation must be used for formative as well as summative program evaluation. In fact, Fixsen, et al., (2005) states that evaluations of newly implemented programs may result in poor results, not because the program at an implementation site is ineffective, but because the results at the implementation site were assessed before the program was completely implemented and fully operational. Measures on implementation must be sensitive enough to measure smaller changes in implementation progress while maintaining a low response cost for administration. Systems that utilize desktop or web-based collection appear promising areas for research and development.
Finally, the study of how context (variables and factors at the classroom, school, and district levels) impact the adoption, installation, and sustainability of educational interventions should continue. Accounting for additional variables (e.g. additional program interventions or emphasis at a school level) is an ongoing area and consideration in implementation research. It is unlikely that any school has only one intervention/program in effect in a given time period. The pursuit of documenting how treatment and programs interact to support or hinder implementation are important areas for future research. The use of statistical tools such as multilevel modeling will allow us to better identify these possibilities and help improve our understanding of complex individual and program outcomes.
REFERENCES


APPENDIX A: LIST OF ACRONYMS/ABBREVIATIONS USED IN CURRENT RESEARCH

- **BOQ**: (Kincaid, Childs, & George, 2005). This instrument is a rating scale containing 53 items, organized into 10 principal components or critical elements of PBIS implementation. This self-assessment is completed by school teams on a yearly basis to assess how they score with regard to developing and implementing school wide PBIS. These scores guide each school’s action plan for the following year.

- **Grandmean**: To fit the Grand Mean centered model, the approach taken was to adjust grade 4 means for each school for the performance of their students as 3\(^{rd}\) graders. The same approach was taken to adjust the grade 5 means for each school for the performance as 4\(^{th}\) graders. Grand mean centering was done with the reading and math scaled scores for each dataset containing the students matched across the 2 years containing their school SET scores. In the first dataset from 2007 to 2008 matched 3\(^{rd}\) and 4\(^{th}\) graders the grand mean centering took into account each student’s 3\(^{rd}\) grade reading scaled score from which the 2007 mean for all 3\(^{rd}\) graders across the whole data set was then subtracted. This was also done for math, ELA, science, and social studies scaled scores. The additional data set from 2008-2009 also had scores computed for each student for grand mean centering all 5 scaled scores.

- **iLEAP**: *Integrated Louisiana Educational Assessment Program*: According to the Louisiana Department of Education (2009), the iLEAP is a criterion-referenced test administered to students in third, fifth, sixth, seventh, and ninth grade in the state of Louisiana. The iLEAP assigns students into one of five categories based on their performance: unsatisfactory, approaching basic, basic, mastery, and advanced. The ELA test consists of four subtests, reading, language, writing, and using information resources. The subtests include 78 multiple-choice items and a writing prompt. The reliability of the third grade ELA test is adequate, with Cronbach’s alpha and the Stratified alpha both revealing a .93 reliability coefficient. The English Language Arts (ELA) and Math portions of the LEAP test are required for promotion from 4\(^{th}\) to 5\(^{th}\) grade. In order to pass to the next grade, students must achieve a combination of at least *Approaching Basic* level on one part of the exam and at least *Basic* on the other portion of the test (LEAP and GEE Interpretive Guide, Louisiana Department of Education, 2009).

- **ISE**: *In School Expulsions*: The rate of students per 100 students where the disciplinary action is exclusion inside of the school building of a student from the student’s regular education program for more than 10 school days for disciplinary reasons by the school principal. There is usually a hearing examiner involved in expulsions.
  - **Change ISE**: Difference between the 2007 and 2008 ISE rates for a school.
**ISS: In School Suspensions:** The rate of students per 100 students where the disciplinary action is exclusion within the school building of a student from the student’s regular education program for up to but not more than 10 school days for disciplinary reasons by the school principal.

- **Change ISS:** Difference between the 2007 and 2008 ISS rates for a school.

**LEAP: Louisiana Education Assessment Program for the 21st Century:** According to the Louisiana Department of Education (2009), LEAP 21 is the state sanctioned criterion-referenced assessments used to monitor how well students have learned the state content standards up to fourth and eighth grades. The purpose of LEAP 21 was therefore, to ensure that fourth and eighth grade students had the knowledge to pass onto the next level of education (i.e., middle school and high school). To determine this, students are given achievement ratings; advanced, mastery, basic, approaching basic and unsatisfactory. In order for students to move to the next grade, they must have and achievement rating of approaching basic or above.

**OSE: Out of School Expulsions:** The rate of students per 100 students where the disciplinary action is exclusion outside of the school building of a student from the student’s regular education program for more than 10 school days for disciplinary reasons by the school principal. There is usually a hearing examiner involved in expulsions.

- **Change OSE:** Difference between the 2007 and 2008 OSE rates for a school.

**OSS: Out of School Suspensions:** The rate of students per 100 students where the disciplinary action is exclusion outside of the school building of a student from the student’s regular education program for up to but not more than 10 school days for disciplinary reasons by the school principal.

- **Change OSS:** Difference between the 2007 and 2008 OSS rates for a school.

**PBIS: Positive Behavior Interventions and Support** defined by the Office of Special Education Programs (OSEP), Technical Assistance Center on Positive Behavioral Interventions and Supports (2009) as “an application of a behaviorally based systems approach to enhance the capacity of schools, families, and communities to design effective environments that improve the fit or link between research validated practices and the environment in which teaching and learning occur.”

**RTI: Response to Intervention:** Refers to academic systems that are grounded in differentiated instruction. Each approach has critical factors and components to be in place at the universal (Tier 1), targeted group (Tier 2), and individual (Tier 3) levels.
- **SET:** (Sugai, Lewis-Palmer, Todd, & Horner, 2001). The SET involves a 2 to 3 hour site visit to a school by an external evaluator. During the site visit, there are observations and interviews with administration, staff, and students. Seven primary feature of universal or school-wide PBIS are measured including the extent to which schools have developed and implemented rules, consistently taught and reinforced rules and expectations, consistently responded to behavior problems, and used teamwork, leadership, and data for decision-making. SET outcomes have been found to be reliable and valid measures of implementation at the universal level and are closely linked to other indicators of school climate and safety (Horner et al., 2004).
VITA

Michelle Lynn Farnsworth Botos is the oldest daughter of Clark and Myra Farnsworth. She was born in Opelousas, Louisiana. However, most of her childhood and adult years were spent in Baton Rouge, Louisiana. She earned her Bachelor of Arts in Speech-Language Pathology from Louisiana Tech University in 2001 and her Master of Arts in Teaching from Southeastern Louisiana University in 2005. Michelle joined Louisiana State University for graduate studies in 2005 and was a graduate assistant working with Positive Behavior Support. She became a coordinator for Positive Behavior Support in 2009. Currently, she is a special education teacher in Livingston Parish. She is married and is the proud mother of four precious children. The degree of Doctor of Philosophy will be conferred by Louisiana State University at the May 2013 Commencement Ceremony.