AN EXAMINATION OF THE RELATIONSHIP BETWEEN INTEGRATED
SYSTEMS MODEL IMPLEMENTATION AND STUDENT OUTCOMES

A dissertation submitted to the
Kent State University College and Graduate School
of Education, Health, and Human Services
in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

by
Amity L. Noltemeyer
August, 2009
A dissertation written by

Amity L. Noltemeyer
B.S., Xavier University, 2002
M.S., Miami University, 2003
Ed.S., Miami University, 2005
Ph.D., Kent State University, 2009

Approved by

_________________________, Co-director, Doctoral Dissertation Committee
Frank J. Sansosti

_________________________, Co-director, Doctoral Dissertation Committee
Caven S. Mcloughlin

_________________________, Member, Doctoral Dissertation Committee
Christopher Was

Accepted by

_________________________, Interim School Director, School of Lifespan
Mary Dellmann-Jenkins Development and Educational Services

_________________________, Dean, College and Graduate School of
Daniel F. Mahony Education, Health, and Human Services
The purpose of this study was to examine the relationship between Integrated Systems Model (ISM) implementation and student outcomes. Specifically, the study examined the effect on academic outcomes of (a) overall ISM implementation level, (b) implementation level of each of the six Key Features of ISM, and (c) implementation level of the academic and behavioral components of ISM. Although the effect of ISM implementation level of behavioral outcomes was also of interest, limited available data precluded appropriate analysis of this issue.

Data were accessed on a sample of 2,660 students attending six Ohio elementary schools. Dependent variables included z-scores for student oral reading fluency assessments. Independent variables included (a) school implementation level of nine components of ISM as assessed by the Implementation Evaluation Tool (IET; State Improvement Grant State Steering Committee, 2007), and (b) initial student reading performance as indicated by assessment scores from one year prior.

Multiple two-way factorial ANOVAs were conducted using IET implementation level and initial student reading performance as independent variables. Results
suggested a main effect of implementation level on student outcomes for (a) overall ISM implementation level, (b) implementation level of four of the six Key Features of ISM, and (c) implementation level of academic components of ISM. Additionally, interaction effects were found across most of the analyses, suggesting a stronger positive effect of implementation level on student academic outcomes for initially Low performing students. The magnitude of the observed effects varied; however, all effects were in the small or medium range.

This study has limitations in its design, internal validity, and sampling. Despite these limitations, several implications appear warranted. First, it is important for schools to monitor implementation levels of ISM and its Key Features. Additionally, results indicate components of ISM that may warrant increased focus during initial implementation efforts. Finally, the study revealed flaws with previous research and directions for future research.
ACKNOWLEDGEMENTS

I would like to acknowledge the efforts of many individuals who contributed to the completion of this milestone. First, I would like to extend my deep appreciation to my committee members. Dr. Frank Sansosti, Dr. Caven Mcloughlin, and Dr. Christopher Was each provided valuable guidance and feedback to enhance the quality of this study. In particular, I am appreciative of the efforts of Dr. Frank Sansosti who worked with me tirelessly from the earliest stages of the process, providing ongoing support and feedback.

Second, I would like to thank other individuals who contributed to my professional development throughout the years. Dr. Katherine Wickstrom helped inspire my initial interest in this topic through her teaching and thesis advising. This interest continued to be fostered by my internship supervisors and co-workers. Dr. Cathy Telzrow developed this interest a step further by providing me with research mentorship. And Shannon Goss, a fellow doctoral candidate, brainstormed ideas and provided support along the way.

Additionally, there are many people who were instrumental in obtaining access to the data used in the study. Most notably, Dr. Karen Schaeffer facilitated the data collection process and educated me on the IET. In addition, Dr. Karen Stine, the building/district representatives, and several individuals from the State Support Team 13 and the University of Cincinnati’s School Psychology program provided assistance along the way.

Finally, I cannot overstate my gratitude to my family for supporting me along this wonderful journey. I want to thank my parents for all they have done for me throughout
the years. Paul, thank you for the countless sacrifices you have made to help me achieve this milestone. You recognized the importance of this endeavor and never once complained about the time, money, or effort devoted. And most of all, I want to thank my beautiful son Pierce for teaching me more than I could ever learn in a class, making me laugh often, and inspiring me to be a better person.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER I: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of Problem</td>
<td>7</td>
</tr>
<tr>
<td>Purpose and Significance of Study</td>
<td>8</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER II: REVIEW OF THE LITERATURE</td>
<td>12</td>
</tr>
<tr>
<td>History, Components, and Examples of RTI, SWPBS, and ISM</td>
<td>12</td>
</tr>
<tr>
<td>History</td>
<td>12</td>
</tr>
<tr>
<td>Need to Improve Outcomes</td>
<td>12</td>
</tr>
<tr>
<td>Frustration with Other Models</td>
<td>14</td>
</tr>
<tr>
<td>Legislative and Policy Support</td>
<td>16</td>
</tr>
<tr>
<td>Research Support</td>
<td>18</td>
</tr>
<tr>
<td>Response to Intervention (RTI)</td>
<td>21</td>
</tr>
<tr>
<td>Conceptual Framework</td>
<td>21</td>
</tr>
<tr>
<td>Components</td>
<td>21</td>
</tr>
<tr>
<td>Examples</td>
<td>28</td>
</tr>
<tr>
<td>Schoolwide Positive Behavior Support</td>
<td>31</td>
</tr>
<tr>
<td>Conceptual Framework</td>
<td>31</td>
</tr>
</tbody>
</table>
Design ................................................................. 69

Procedures ............................................................ 70

Data Analysis .......................................................... 71

  Descriptive Statistics ................................................. 71
  Analysis of Covariance ............................................... 72
  Factorial Analysis of Variance ..................................... 73
  Statistical Assumptions ............................................. 74
  Measurement Assumptions ......................................... 77

CHAPTER IV: RESULTS ............................................... 79

  Overview ............................................................ 79

  Descriptive Statistics .............................................. 79

  Research Question One ............................................ 82

    Hypotheses ....................................................... 82
    Findings .......................................................... 83

  Research Question Two ........................................... 86

    Hypotheses ....................................................... 86
    Findings .......................................................... 86

      Administrative Leadership .................................... 87
      Culturally Responsive Practices .......................... 89
      Collaborative Strategic Planning .......................... 91
      Data-based Decision Making ................................. 94
      Scientifically Based Research ............................... 96
Implications for Practice ........................................ 123
Implications for Research ................................. 126
APPENDICES ...................................................... 129
   Appendix A: Implementation Evaluation Tool Phase I ............. 130
   Appendix B: Implementation Evaluation Tool Phase II .............. 134
   Appendix C: Implementation Evaluation Tool Phase III ............. 141
   Appendix D: Consent Letter for Release of Data .................... 149
REFERENCES ...................................................... 154
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Tiered ISM Model (from Graden et al., 2007)</td>
<td>22</td>
</tr>
<tr>
<td>2.</td>
<td>2008 Mean Student DIBELS Z-Scores by School IET Implementation Level</td>
<td>83</td>
</tr>
<tr>
<td>3.</td>
<td>2008 Student DIBELS Z-Scores by School IET Implementation Level and spring 2007 DIBELS Z-Scores</td>
<td>85</td>
</tr>
<tr>
<td>4.</td>
<td>Figure 3 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment</td>
<td>85</td>
</tr>
<tr>
<td>5.</td>
<td>2008 Mean Student DIBELS Z-Scores by School AL Implementation Level</td>
<td>88</td>
</tr>
<tr>
<td>6.</td>
<td>2008 Student DIBELS Z-Scores by School AL Implementation Level and spring 2007 DIBELS Z-Scores</td>
<td>88</td>
</tr>
<tr>
<td>7.</td>
<td>Figure 6 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment</td>
<td>89</td>
</tr>
<tr>
<td>8.</td>
<td>2008 Mean Student DIBELS Z-Scores by School CRP Implementation Level</td>
<td>90</td>
</tr>
<tr>
<td>9.</td>
<td>2008 Student DIBELS Z-Scores by School CRP Implementation Level and spring 2007 DIBELS Z-Scores</td>
<td>90</td>
</tr>
<tr>
<td>10.</td>
<td>Figure 9 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment</td>
<td>91</td>
</tr>
</tbody>
</table>
11. 2008 Mean Student DIBELS Z-Scores by School CSP Implementation Level ........................................ 92
12. 2008 Student DIBELS Z-Scores by School CSP Implementation Level and spring 2007 DIBELS Z-Scores ........................................ 93
13. Figure 12 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment ................. 93
14. 2008 Mean Student DIBELS Z-Scores by School DBDM Implementation Level ........................................ 95
15. 2008 Student DIBELS Z-Scores by School DBDM Implementation Level and spring 2007 DIBELS Z-Scores ........................................ 95
16. Figure 15 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment ................. 96
17. 2008 Mean Student DIBELS Z-Scores by School SBR Implementation Level ........................................ 97
18. 2008 Student DIBELS Z-Scores by School SBR Implementation Level and spring 2007 DIBELS Z-Scores ........................................ 98
19. Figure 18 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment ................. 98
20. 2008 Mean Student DIBELS Z-Scores by School TT Implementation Level ........................................ 100
21. 2008 Student DIBELS Z-Scores by School TT Implementation Level and spring 2007 DIBELS Z-Scores ........................................ 100
22. Figure 21 modified to exclude Medium implementers and Medium
performers on the spring 2007 DIBELS assessment. .................. 101
23. 2008 Mean Student DIBELS Z-Scores by School ACADEM Implementation
Level ................................................................. 104
24. 2008 Student DIBELS Z-Scores by School ACADEM Implementation Level
and spring 2007 DIBELS Z-Scores ................................. 104
25. Figure 24 modified to exclude Medium implementers and Medium
performers on the spring 2007 DIBELS assessment. .............. 105
26. 2008 Mean Student DIBELS Z-Scores by School BEHAV Implementation
Level ................................................................. 106
27. 2008 Student DIBELS Z-Scores by School BEHAV Implementation Level
and spring 2007 DIBELS Z-Scores ................................. 106
28. Figure 27 modified to exclude Medium implementers and Medium
performers on the spring 2007 DIBELS assessment. .............. 107
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Description of the Six Key Features of Ohio’s ISM</td>
<td>41</td>
</tr>
<tr>
<td>2. T-Test ( p )-Values for the Eight Schools on Six Demographic Variables</td>
<td>59</td>
</tr>
<tr>
<td>3. Comparison of Mean Demographic Information for Participating Schools versus Statewide Averages (Ohio Department of Education, 2009)</td>
<td>60</td>
</tr>
<tr>
<td>4. Summary of Variables by Research Question</td>
<td>67</td>
</tr>
<tr>
<td>5. Theoretical Loadings for the Six Scales of the IET</td>
<td>68</td>
</tr>
<tr>
<td>6. Theoretical Loadings for the Six Scales of the IET on the Academic and Behavior Components</td>
<td>69</td>
</tr>
<tr>
<td>7. Kolmogorov-Smirnoff Results by Group</td>
<td>75</td>
</tr>
<tr>
<td>8. Levene Test Results for All Analyses</td>
<td>77</td>
</tr>
</tbody>
</table>
CHAPTER I: INTRODUCTION

As a result of “the widespread public perception that something is seriously remiss in our educational system,” United States Secretary of Education T.H. Bell established the National Commission on Excellence in Education on August 26, 1981 to examine the quality of education in the United States (National Commission on Excellence in Education, 1983, Introduction Section). The report that emerged from this investigation, A Nation at Risk (National Commission on Excellence in Education), highlighted the gross inadequacies of public education in the United States at the time. The Commission reported that 23 million American adults were functionally illiterate, achievement test scores had been declining over the past two decades, and Americans were significantly underperforming other industrialized nations across academic domains. Such subpar academic outcomes were deemed to be a result of considerable deficiencies in curricular content, academic expectations, instructional time, and teacher preparation. Consequently, the Commission recommended the establishment of rigorous academic standards, increased instructional time via lengthened school days and school years, improved teacher preparation programs to enhance the quality of teachers, and interventions to improve student outcomes.

Unfortunately, relatively few of these recommendations were fully embraced, with student performance concerns persisting to this day. For example, recent data suggest only 29% and 32% of all 8th grade students scored at or above the Proficient range on the National Assessment of Education Progress (NAEP) Reading and Mathematics tests, respectively (Snyder, Dillow, & Hoffman, 2008). These figures are
more troubling for African American (12% and 14%, respectively) and low income (15% and 17%, respectively) students (Snyder et al.). Additionally, only 13% of adults scored at or above the Proficient level on basic tests of literacy, suggesting an overwhelming inability to engage in complex reading activities (Snyder et al.). Although these indicators of achievement merely scratch the surface, they provide a glimpse into the continued academic needs of American students.

In addition to these academic needs, the behavioral needs of students – which were not addressed in the groundbreaking Nation at Risk (National Commission on Excellence in Education, 1983) report – have become increasingly problematic for educational professionals. Ysseldyke et al.’s (1997) assertion that, “The student population entering America’s classrooms is more challenging than at any time in our recent history,” has been repeatedly illustrated by increased examples of high profile school-related violent deaths (e.g., Columbine), as well as less severe but higher incidence student and staff victimization. For example, recent data suggest that 14% of students in grades 9-12 reported being in a fight at school in the past year, 6% of students in grades 9-12 reported carrying a weapon to school in the past 30 days, and 28% of students ages 12–18 reported being bullied at school during the past six months (Dinkes, Cataldi, Lin-Kelly, & Snyder, 2007). In addition, 7% of teachers reported having been threatened with physical injury or physically attacked and 8% of students reported being threatened or injured with a weapon during the past year (Dinkes et al.). Such behaviors impede student learning by decreasing instructional time and inhibiting a focus on the learning process.
In response to such startling academic and behavioral trends, the past decade has been characterized by the realization that the recommendations made by the authors of *A Nation at Risk* (1983) nearly two decades earlier desperately needed to be heeded in order to improve outcomes for all students. This realization served as an impetus behind the development of federal legislation such as the No Child Left Behind Act (NCLB, 2001), which requires schools to demonstrate accountability for the outcomes of all students, ensure that students are taught using evidence-based or scientifically-based practices, and provide services to prevent or intervene upon academic problems. The Individuals with Disabilities Education Act (IDEA, 2004), the most recent version of the federal legislation governing services for students with disabilities, also emphasizes enhancing the outcomes of students with disabilities by including them in accountability and assessment systems, requiring pre-referral interventions prior to special education identification, and mandating protections and services for students whose behavioral manifestations of their disability interfere with the learning process.

Currently, students are entering schools with increasingly diverse needs, and, at the same time, teachers are expected to hold them to higher standards than ever before. Given the increased challenges of student diversity, coupled with the failure of traditional models to improve student outcomes, educators have begun searching for new ways to serve students more effectively. Comprehensive School Reform (CSR) models are one increasingly popular vehicle for meeting this end. CSR models have been defined as coherent schoolwide improvements that address all aspects of a school’s operations rather than only fragmented pieces (U.S. Department of Education, 2004). Among other
components, effective CSR models typically integrate: (a) a comprehensive design with aligned components, (b) ongoing, high-quality professional development, (c) measurable goals and benchmarks for student achievement, (d) support for teachers, administrators and staff, (e) parent and community involvement, (f) evaluation of strategies for the implementation of the reform, (g) resources to support and sustain the reform, and (h) demonstrated research effectiveness (U.S. Department of Education, 2004). Two examples of CSR aligned with current legislation that are gaining increasing support are Response to Intervention (RTI) and Schoolwide Positive Behavior Support (SWPBS).

RTI has been defined broadly as, “…the practice of (1) providing high-quality instruction/intervention matched to student needs and (2) using learning rate over time and level of performance to (3) make important education decisions (Batsche et al., 2006). Essentially, it represents a framework for improving academic and/or behavioral outcomes for all students, although the primary focus has been on the former. Grounded in a prevention focus, RTI typically incorporates the following components: a tiered model of support, data collection, data-based decision making, problem-solving, evidence-based instruction, treatment fidelity, and professional development. Using an RTI approach, students are exposed to increasingly intense interventions. Based on their individual response to intervention, as measured by ongoing data collection, important educational decisions are made regarding the effectiveness of instruction. If a student continues to demonstrate limited progress after multiple iterations of individually matched interventions, he or she may be referred for special education services.
SWPBS is a similar educational initiative designed to improve student outcomes. Specifically, SWPBS is a comprehensive approach designed to promote the appropriate behaviors of all students and enhance the capacity of schools and families to design positive environments for students (OSEP Center on Positive Behavior Interventions and Supports, 2004). SWPBS has been defined as, “…the application of positive behavioral interventions and systems to achieve socially important behavior change” (Sugai, Horner, et al., 2000, p.133). Similar to RTI, SWPBS involves a tiered system of prevention and intervention and an emphasis on professional development, treatment integrity, evidence-based instruction, and data-based decision-making. For example, in a SWPBS model, evidence based instruction is provided at the schoolwide (Tier I), classroom (Tier II), and individual (Tier III) level based on needs revealed by data. SWPBS typically involves six components: (a) identifying a statement of purpose, (b) establishing schoolwide behavioral expectations, (c) teaching schoolwide expectations, (d) encouraging expected behaviors, (e) discouraging problem behaviors, and (f) engaging in data collection and decision making (Lewis & Sugai, 1999).

Given the shared features of RTI and SWPBS, researchers and educational professionals have increasingly been interested in the promise of streamlining the process by incorporating both initiatives into a comprehensive school program. This interest has been fueled by research demonstrating causal paths between academic and behavior outcomes for students (e.g., Kellam, Mayer, Rebok, & Hawkins, 1998). Often referred to as an Integrated Systems Model (ISM), this type of model involves implementing tiered models of academic (RTI) and behavioral (SWPBS) support in a simultaneous and
streamlined fashion within a school, district, or state. Of particular interest in the present study is Ohio’s Integrated Systems Model (referred to as OISM for clarity, although it is now often referred to as ISM and many components of OISM are incorporated into the more recently developed Ohio Improvement Process). OISM is a prevention-based model designed to improve academic and behavioral outcomes at multiple levels (e.g., district, school, classroom, individual student) (Graden, Stollar, & Poth, 2007). It is based on six Key Features: (a) Collaborative Strategic Planning, (b) Data-based Decision Making, (c) Academic and Behavioral Supports across Three Tiers, (d) Scientifically-based Practices, (e) Administrative Leadership, and (f) Culturally Responsive Practices (Graden et al.; Stollar, Poth, Curtis, & Cohen, 2006). Both building-level and district-level teams are used to facilitate the design, implementation, and evaluation of ISM within a school district (Graden et al.).

Research has demonstrated positive student outcomes associated with both large-(i.e., statewide, districtwide, schoolwide) and small-scale implementation (i.e., classwide, small group, individual student) of RTI (e.g., Burns, Appleton, and Stehouwer, 2005) and SWPBS (e.g., Barrett, Bradshaw, & Lewis-Palmer, 2008). In addition, although limited in number and scope, initial findings suggest improved outcomes associated with ISMs surpass those found with either RTI or SWPBS in isolation (e.g., Stewart, Benner, Martella, & Marchand-Martella, 2007).

Despite these promising results, many schools have struggled with implementation of these initiatives. Because RTI, SWPBS, and ISM represent comprehensive educational reform, they require vast shifts in the organization,
leadership, roles, practices, and beliefs of educators at all levels. Unfortunately, previous research has shown that implementation of reform initiatives has proven difficult (e.g., Berends, Bodilly, & Kirby, 2002; Mann, 1978) and low implementation levels have resulted in disappointing outcomes (e.g., Datnow, Borman, & Stringfield, 2000; Nunnery, 1997). Although some studies on RTI, SWPBS, and ISM have included implementation data as a supplement to their research questions, there are no known studies that have systematically verified this relationship between quality of implementation and student outcomes for these specific initiatives.

Statement of the Problem

Despite the promise of ISM for improving student academic and behavioral outcomes, there remain several unanswered questions regarding implementation and utility. First, preliminary evidence exploring small-scale application suggests that ISMs may be associated with improved student academic and behavioral outcomes. However, these findings have not been convincingly extended to include large-scale models over an extended period of time. Given the time, resources, and intensity required to implement large-scale educational change such as ISM, it is important to document the extent to which student outcomes are improved as a result.

Second, there has been little research examining the implementation of ISM initiatives. Given the difficulties schools often encounter when implementing comprehensive changes such as ISM, it is important to determine whether or not degree of implementation is related to degree of improvement in student outcomes. Furthermore, it is important to identify those specific components of ISM that have the
largest influence on student outcomes. These keystone components can then serve as the initial areas for schools to focus on when attempting to improve implementation.

Purpose and Significance of Study

The purpose of this study is to explore the relationship between: (a) ISM implementation and student behavioral and academic outcomes, (b) implementation of each of the six Key Features of ISM and student academic outcomes, and (c) implementation of the academic/behavioral components of ISM and student academic outcomes.

The significance of this study can be viewed in terms of both research and practical application. This study will contribute to the extant research literature in three ways. First, it will represent one of a small number of studies examining student behavioral and academic outcomes related to large-scale implementation of ISM. Second, this study will represent the first known attempt to systematically examine the relationship between degree of ISM implementation and student outcomes. Third, it will be the first known project to identify key components of ISM most significantly related to improvements in student outcomes. It is anticipated that results of this study can be used to inform school-based practice by demonstrating the outcomes of ISM and identifying the key components of ISM on which to focus initial implementation efforts.

Research Questions and Hypotheses

This study is designed to address four research questions. Each of these will be presented, followed by a hypothesis for each:
**Research Question One:** Are there significant differences in student academic outcomes based on school ISM implementation level?

Hypothesis: Significant differences in student academic outcomes were expected based on school ISM implementation level. Specifically, High implementing schools were expected to have higher student academic outcomes than Medium and Low implementing schools. In addition, Medium implementing schools were expected to have higher student academic outcomes than Low implementing schools.

**Research Question Two:** Are there significant differences in student academic outcomes based on school implementation level of each of the Six Key Features of Ohio’s ISM?

Hypothesis: It was expected that differences in student academic outcomes would be found based on school implementation level of each of the Six Key Features of Ohio’s ISM. For each feature, it was anticipated that High implementing schools would be found to have higher student academic outcomes than Medium and Low implementing schools. In addition, Medium implementing schools would be found to have higher student academic outcomes than Low implementing schools. However, it was anticipated that differences in the strength of the relationship between implementation level and student academic outcomes would be found. For example, it was anticipated that the proportion of variance in student academic outcomes explained by each of the Six Key Features would differ, speaking to the relative importance of each factor related to student outcomes. Given the lack of previous research exploring the relative importance of each Key Feature, there were no hypotheses regarding which would be most significant.
**Research Question Three:** Are there significant differences in student academic outcomes based on school implementation level of the (a) academic components of ISM and (b) behavioral components of ISM?

Hypothesis: It was expected that differences in student academic outcomes would be found based on school implementation level of the academic and behavioral components. For both components, it was anticipated that High implementing schools would be found to have higher student academic outcomes than Medium and Low implementing schools. In addition, Medium implementing schools would be found to have higher student academic outcomes than Low implementing schools. However, it was anticipated that differences in the strength of the relationship between implementation level and student academic outcomes would be found. Specifically, it was anticipated that the proportion of variance in student academic outcomes explained by the academic and behavior components would differ, with a greater proportion of variance explained by the academic components. In addition, it was expected that the proportion of variance explained by the integrated model (research question 1) would be greater than that explained by either the academic or behavioral components in isolation.

**Research Question Four:** Are there significant differences in school-level behavioral outcomes based on school-level ISM implementation level?

Hypothesis: It was anticipated that significant differences in school-level behavioral outcomes would be found based on school-level ISM implementation level. Specifically, High implementing schools would be found to have more favorable school-level behavioral outcomes than Medium implementing schools and Medium
implementing schools would have more favorable school-level behavioral outcomes than Low implementing schools.
CHAPTER II: REVIEW OF THE LITERATURE

History, Components, and Examples of RTI, SWPBS, and ISM

As a result of their potential to improve student outcomes, RTI, SWPBS, and ISM are increasingly being adopted nationwide. This section will review (a) the history of RTI, SWPBS, and ISM, (b) the conceptual framework, components, and field examples of RTI, SWPBS, and ISM, (c) the student academic and behavior outcomes associated with RTI, SWPBS, and ISM, and (d) the relationship between implementation of other school reform initiatives and student outcomes. Because the focus of this investigation is on the implementation of a statewide initiative, preference will be given to large-scale models over small-scale models; however, research on the latter will be used in the absence of former.

History

Many forces worked in conjunction to bring about the emergence of RTI, SWPBS, and ISM. These included (a) an identified need to improve student outcomes, (b) frustration with other models for improving student outcomes, (c) legislative and policy support, and (d) results of research on learning/intervention processes. In the following sections, a brief history of these factors will be explored in order to provide a foundation for understanding current models and practices.

Need to Improve Outcomes

As a result of “the widespread public perception that something is seriously remiss in our educational system,” United States Secretary of Education T.H. Bell established the National Commission on Excellence in Education on August 26, 1981 to
examine the quality of education in the United States. The report that emerged from this investigation, *A Nation at Risk* (National Commission on Excellence in Education, 1983), highlighted the gross inadequacies of public education in the United States at the time. The Commission reported that 23 million American adults were functionally illiterate, achievement test scores had been declining over the past two decades, and Americans were significantly underperforming other industrialized nations across academic domains. Such subpar academic outcomes were deemed to be a result of considerable deficiencies in curricular content, academic expectations, instructional time, and teacher preparation. Consequently, the Commission recommended the establishment of rigorous academic standards, increased instructional time via lengthened school days and school years, improved teacher preparation programs to enhance the quality of teachers, and interventions to improve student outcomes.

Unfortunately, relatively few of these recommendations were fully embraced, with student performance concerns persisting to this day. For example, recent data suggest only 29% and 32% of all 8th grade students scored at or above the Proficient range on the National Assessment of Education Progress (NAEP) Reading and Mathematics tests, respectively (Snyder et al., 2007). These figures are more troubling for African American (12% and 14%, respectively) and low income (15% and 17%, respectively) students (Snyder et al.). Additionally, only 13% of adults scored at or above the Proficient level on basic tests of literacy, suggesting an overwhelming inability to engage in complex reading activities (Snyder et al.). Although these indicators of
achievement merely scratch the surface, they provide a glimpse into the continued
academic needs of American students.

In addition to these academic needs, the behavioral needs of students – which
were not addressed in the groundbreaking Nation at Risk (National Commission on
Excellence in Education, 1983) report – have become increasingly problematic for
educational professionals. Ysseldyke et al.’s (1997) assertion that, “The student
population entering America’s classrooms is more challenging than at any time in our
recent history,” has been repeatedly illustrated by the increased examples of high profile
school-related violent deaths (e.g., Columbine) as well as less severe but higher
incidence forms of student and staff victimization. Recent data suggest that 14% of
students in grades 9-12 reported being in a fight at school in the past year, 6% of students
in grades 9-12 reported carrying a weapon to school in the past 30 days, 28% of students
ages 12–18 reported being bullied at school during the past six months, 7% of teachers
reported having been threatened with physical injury or physically attacked during the
past year, and 8% of students reported being threatened or injured with a weapon during
the past year (Dinkes et al., 2007). Such behaviors impede student learning by
decreasing instructional time and inhibiting a focus on the learning process.

Frustration with Other Models

Together, these academic and behavioral data highlight the ongoing need to
revolutionize the way schools operate in order to enhance student outcomes. Early on,
special education programming was perceived as a model for improving the outcomes of
struggling learners. However, research over the years has repeatedly questioned the
effectiveness of special education at meeting this end (e.g., Cegelka & Tyler, 1970; Carlberg & Kavale, 1980). In addition, dissatisfaction with special education identification processes (e.g., the discrepancy model for identifying learning disabilities) and concerns over the rapidly increasing number of students identified as needing special education services were cause for concern.

In an attempt to address these issues, “prereferral” intervention teams emerged in the 1980’s (Fuchs, Mock, Morgan, & Young, 2003). These collaborative teams (also known as Student Study Teams, Instructional Support Teams, Intervention Assistance Teams, and Teacher Assistance Teams) sought to improve student outcomes through individual problem-solving while concurrently reducing the number of unnecessary special education evaluations. Specifically, if a teacher had concerns about a student’s behavioral or academic functioning, the team would meet with the student’s teacher and parent in order to engage in collaborative problem-solving prior to referring the student for a special education evaluation (Gersten & Dimino, 2006). This collaborative problem-solving process involved four steps repeated in a cyclical fashion: (a) problem identification, (b) problem analysis, (c) plan implementation, and (d) plan evaluation (Allen & Graden, 2002). The prereferral intervention team concept spread quickly, and by 1989 it was required in 23 states and recommended in 11 others (Carter & Sugai, 1989).

Although some benefits were reaped from prereferral intervention teams, they were often viewed as a meaningless additional procedure for requesting a traditional special education evaluation and were not highly valued by teachers (Gersten & Dimino,
The focus was not on improving outcomes for students by developing and implementing quality interventions, but rather on determining whether or not the child should be referred for special education services. In addition, teachers often sought prereferral intervention only when they already thought the student needed help beyond the help they could provide, resulting in interventions that often backfired (Gersten & Dimino). Finally, because problem-solving was done on an individual student basis, and the demand of students with concerns outnumbered the ability of staff to perform individual problem-solving and intervention, the model lacked efficiency.

RTI and SWPBS slowly emerged as attempts to improve upon the failure of both special education and pre-referral teams for improving outcomes for all students. Specifically, the focus of these models is more preventative, systemic, and outcomes-oriented. By using a tiered system, interventions can be provided at the classwide, small group, and individualized level, thus maximizing resources and facilitating interventions based on student need. Although RTI was originally recognized as an alternative method for preventing unnecessary special education referrals and determining special education eligibility, its current conceptualization has now aligned with SWPBS and ISM to ensure improved outcomes for all students.

Legislative and Policy Support

Another factor facilitating acceptance of RTI, SWPBS, and ISM was increasing legislative and policy support. A realization that the accountability recommendations of A Nation at Risk (1983) desperately needed to be heeded served as an impetus behind the development of federal legislation such as the No Child Left Behind Act (NCLB, 2001).
NCLB requires schools to demonstrate accountability for the outcomes of all students, ensure that students are taught using evidence-based or scientifically-based practices, and provide services to prevent or intervene upon academic and behavioral problems. In addition, under NCLB, school districts are required to demonstrate an annual increase in the percentage of students demonstrating proficiency on accountability measures until 100% of students (including students with disabilities) are proficient by 2014. Finally, NCLB also addresses the importance of creating a positive school climate. For example, it stipulates that parents of children attending “persistently dangerous” schools can transfer their children to another school in the district. These stipulations are aligned closely with the goals of RTI, SWPBS, and ISM.

The Individuals with Disabilities Education Act (IDEA, 2004), the federal legislation governing special education, also has contributed to the growth of RTI, SWPBS, and ISM. For example, IDEA permits states to use a child’s response to intervention in determining whether or not that child qualifies as a student with a specific learning disability. In addition, the law requires districts to provide interventions for students and ensure a quality curriculum was provided prior to conducting a multifactored evaluation and mandates that teams consider using positive behavioral support when a student’s behavior interferes with learning. IDEA also allows school districts to use up to 15% of their Part B funds for early intervention services, which can be used to support RTI and SWPBS processes. Finally, the educational team is required to conduct a functional behavior assessment and behavioral intervention plan within 10 days after a student with a disability has been disciplined.
In addition to legislation such as NCLB (2001) and IDEA (2004), numerous educational organizations have expressed their support for these models through policy analyses. For example, the National Joint Council on Learning Disabilities (NJCLD, 2005) identified RTI as a promising method and the President’s Commission on Excellence in Special Education (2002) called for a focus on outcomes and prevention, consistent with an RTI approach. In addition, the National Association of State Directors of Special Education (Batsche et al., 2006) published policy considerations supporting RTI, which are widely used in the field. SWPBS has also been well supported by organizations. For example, the Association for Positive Behavior Support is a multidisciplinary international organization seeking to expand the application of PBS strategies (Association for Positive Behavior Support, 2007). In addition, the National Technical Assistance Center on Positive Behavior and Intervention Supports (PBIS; http://www.pbis.org) provides schools with capacity-building information and technical support to create more effective schools through SWPBS implementation.

Research Support

An additional factor contributing to the development of RTI, SWPBS, and ISM was the advances in research that supported these models. For example, several lines of research on assessment and intervention have contributed to the refinement of RTI processes. First, because data collection and data-based decision-making are essential components of RTI, the development of curriculum-based measurement (CBM) tools for screening and progress monitoring has been highly influential. CBM is an assessment tool that involves the frequent administration of brief, timed skill probes that assess both
fluency and accuracy (Shapiro, 1996). Without the pioneering work of Deno & Mirkin (1977) and Deno (1985), and the plethora of subsequent studies validating and expanding the use of CBM, RTI in its current form would not be possible. Although CBM began in the area of basic reading skills (e.g., fluency), it has now expanded to included validated measures in math, spelling, writing, reading comprehension, and other content areas (Hosp, Hosp, & Howell, 2007). In recent years, these assessments have become quite user-friendly. In fact, consumers can download and print Dynamic Indicators of Basic Literacy Skills (DIBELS; Good & Kaminski, 2002) reading measures for no cost.

In addition, research on effective instruction and intervention programs facilitated the development of RTI. Because a critical component of RTI is evidence-based or scientifically-based practice, RTI became more feasible as an increasing number of studies have verified the utility of intervention programs at each of the three Tiers. In addition to several reputable resources summarizing the research support for various intervention programs (e.g., Florida Center for Reading Research website, What Works Clearinghouse website), there have been a plethora of user-friendly websites designed to share research-based interventions with practitioners (e.g., interventioncentral.org).

Similar research helped propel SWPBS into action. SWPBS was originally derived as an alternative to punitive and aversive interventions used to control the behavior of students with aggressive or self-injurious behavior (Hieneman, Dunlap, & Kincaid, 2005). Many of the foundational components of SWPBS were derived from Applied Behavior Analysis (ABA), a research-based practice in which learning principles are applied to affect meaningful change in behavior (Dunlap, 2006). In 1991, Colvin,
Sugai, and Kame’enui began a project looking at incorporating features of ABA into proactive school-wide discipline systems (Project PREPARE, OSEP, U.S. Department of Education, Grant No. HO29K10092). They concluded that effective, proactive, scientifically-supported behavior interventions could be applied at the whole-school level (Colvin, Sugai, & Kame’enui, 1994). Several years later, based on the U.S. Department of Health’s tiered model of prevention, Walker and Horner (1996) went beyond school-wide discipline applications to propose a three-tiered model for meeting the behavioral needs of students.

Based on the findings of Colvin et al. (1994), coupled with the framework proposed by Walker and Horner (1996), researchers began incorporating many of the components of PBS into a continuum of services known as SWPBS. By 1998, these efforts led to the emergence of the national SWPBS initiative. SWPBS gained such notice that the Journal of Positive Behavior Interventions was established in 1999 to disseminate research and information on the topic (Dunlap & Koegel, 1999). Research published in this and related journals has supported the continued development and refinement of the model.

It is clear that multiple factors contributed to the development of RTI and SWPBS. Because ISM is based on RTI and SWPBS models, its history is inextricably intertwined with the history of these two initiatives. As each initiative gained momentum and research support, schools, districts, and states began to realize the common features among them (e.g., data-based decision making, teaming and collaboration, tiered model, prevention focus, etc.). In addition, research documenting causal paths between
academic and behavioral outcomes (e.g., Kellam et al., 1998) strengthened the desire to address both. Soon, both school-based and state-initiated integrated models of behavior and academic support emerged.

Response to Intervention (RTI)

Conceptual Framework

RTI has been defined broadly as, “…the practice of (1) providing high-quality instruction/intervention matched to student needs and (2) using learning rate over time and level of performance to (3) make important education decisions (Batsche et al., 2006). Essentially, it represents a framework for improving academic and/or behavioral outcomes for all students, although the primary focus has been on the former. Using an RTI approach, students are exposed to increasingly intense interventions in the area of concern based on their individual response to intervention as measured by ongoing data collection. When students continue to fail, important educational decisions are made regarding the effectiveness of instruction and they may be referred for special education services.

Components

Because RTI is a comprehensive reform model rather than a specific curriculum, variations in the components exist from setting to setting. However, there are several components that are relatively standard and warrant further discussion.

Tiers of support. RTI consists of a continuum of support provided through a three-tiered model (see the left side of the cone in Figure 1) based on a public health model of prevention and intervention (Mellard & Johnson, 2008). The basic premise
behind the model is that each student receives the intensity of support necessary for his or her particular concerns. In Tier I of the RTI model, all students are provided quality research-based instruction in the general education classroom (Mellard & Johnson). With quality Tier I instruction in place, it is expected that approximately 80% of students will meet expectations. The 20% of students whose skill level and rate of progress remain behind peers despite Tier I instruction begin to receive supplemental research-based small group intervention at Tier II (Mellard & Johnson). Finally, approximately 5% of students may continue to lack appropriate progress after Tier I and Tier II interventions.
(Mellard & Johnson). These students then receive intensive individualized research-based interventions at Tier III. Tier III may or may not result in a referral for special education determination (Mellard & Johnson).

Data collection/measurement. Another component of RTI is frequent data collection and monitoring of progress. Within each tier, student progress is typically monitored using curriculum-based measurement (CBM) data. CBM is the preferred method because it is sensitive to change, time/cost efficient, and linked to the curriculum. CBM has been validated for use as a screening and progress monitoring tool in a variety of studies (e.g., Shinn, 1989, 1998).

At Tier I, all students are administered universal screening assessments in order to determine which students may need additional intervention (Mellard & Johnson, 2008). Typically, this involves administering CBM assessments to all students three or four times per year. At Tiers II and III, students are administered progress monitoring assessments to determine the effectiveness of interventions and make changes as needed (Mellard & Johnson). At Tier II, students are typically administered CBM assessments at least twice monthly, and at Tier III students are typically administered the assessments at least weekly. However, variations of the frequency of assessments are determined based upon individual student need.

In addition to student-level data, systems-level data should also be collected. For example, data on schoolwide needs, resources, implementation levels, and curricula can be used to guide RTI development and evaluation. Regarding the latter, the Planning and Evaluation Tool for Effective Schoolwide Reading Programs- Revised (PET-R;
Kame‘enui & Simmons, 2003) is an assessment that can be used to internally or externally evaluate a school’s Tier I reading curricula to determine which components of effective instruction are in place. Results of the assessment can be used by the school to target areas for improvement and monitor progress over time.

**Data-based decision making.** In an RTI model, decisions are made based on the previously mentioned data. It is important to use these data to inform decision-making at both the individual student and systems level. One set of procedures for making data-based decisions at the student level is the use of a dual discrepancy criterion. A student is designated as dually discrepant when he or she performs below peer norms in both level of performance and rate of progress in a particular academic skill area (Fuchs & Fuchs, 1998). Dual discrepancy status has been used to identify students who (a) need additional intervention (e.g., Tier II or Tier III), and/or (b) may have a specific learning disability as evidenced by their non-responsiveness to additional interventions. CBM progress monitoring data are typically used as a basis for designating dual discrepancy status. Researchers have found preliminary evidence for the validity of the dual discrepancy model in identifying students with significant reading concerns (e.g., Speece, Case, & Molloy, 2003; Speece & Case, 2001; McMaster, Fuchs, Fuchs, & Compton, 2005). Although dual discrepancy is one method for making data-based decisions, there are other methods as well. Ultimately, schools can determine their own criteria based on local needs.

**Problem-solving process.** Problem-solving is a generic term used to describe a process by which individual student concerns are identified, analyzed, intervened upon,
and then evaluated. The collaborative problem-solving process is one such framework for addressing student concerns (Allen & Graden, 2002). This problem-solving framework includes four steps: (a) problem identification, (b) problem analysis, (c) plan implementation, and (d) plan evaluation (Allen & Graden). These steps are often repeated as necessary in a cyclical process until effective interventions are found for each student. Although there are other slightly different approaches to problem-solving (e.g., Deno, 2005), each shares a structured, step-like, cyclical process.

Although problem-solving models were traditionally designed to promote individual student change, more recent models have emerged that also address system-level change. For example, Collaborative Strategic Planning (CSP) was designed to address concerns at the student, school, district, region, or state level (Stollar et al., 2006). This model recognizes that for some concerns, problem-solving on an individual level may burden the resources of schools (e.g., personnel time spent attending individual problem-solving meetings and implementing individualized interventions) and a systemic approach is more cost and resource efficient (Stollar et al.). CSP involves two stages. At the first stage, the educational team engages in the following six-step system-level problem-solving process at least annually: (a) system-level problem identification based on student outcome data, (b) system-level problem analysis based on team surveys, (c) identification of priority strengths and needs, (d) system-level goal setting, (e) system-level plan development and implementation, and (f) system-level plan evaluation. In the second stage of CSP, the following five-step problem-solving issue is used to address a specific issue identified in Stage 1 of CSP: (a) problem identification, (b) problem
analysis, (c) goal setting, (d) plan development and implementation, (e) plan evaluation. These processes can be used to facilitate systems change. Whereas the individual problem-solving process may target the causes and interventions for an individual student’s reading fluency deficit, CSP may target the causes and interventions for an entire grade level that in which the majority of student’s failed to meet benchmark expectations on a reading fluency screening assessment.

*Evidence-based interventions and curriculum.* It is important to use evidence-based instruction and intervention as part of an RTI model in order to increase the probability of positive student outcomes (Brown-Chidsey & Steege, 2005). Instruction and/or curricula are considered to be evidence-based when documented to have positive outcomes using a sound experimental design (Brown-Chidsey & Steege). Given wide variations in what is assumed to be “sound experimental design,” a task force on evidence-based interventions sponsored by Division 16 of the American Psychological Association (APA) developed a Procedural and Coding Manual for Review of Evidence-Based Interventions to assist in determining whether interventions are evidence-based (Kratochwill & Stoiber, 2002). This manual provides guidelines for examining the measurement, comparison group, primary/secondary outcomes, educational/clinical significance, durability of effects, intervention components, and treatment fidelity of available intervention programs (Kratochwill & Stoiber). Based on the APA’s guidelines, an intervention must have been supported by at least one between-group or single-subject study and meet certain minimum scores on the aforementioned categories in order to be designated as evidence-based (Kratochwill & Stoiber).
Other resources are also available to help educators identify and select research-based instruction and intervention programs. For example, the What Works Clearinghouse (www.w-w-c.org), the Promising Practices Network (www.promisingpractices.net), and the Florida Center for Reading Research (www.fcrr.org) all assess and summarize the effectiveness research on instruction and intervention programs to help educators make informed decisions. In addition, the U.S. Department of Education (2003) has published a user-friendly guide to identifying and evaluating research-based interventions.

*Treatment integrity.* Treatment integrity (aka treatment fidelity, intervention fidelity, and intervention integrity) describes the degree to which an intervention is implemented as designed or intended (Gresham, 1989; Gresham, Gansle, & Noell, 1993; Gresham, MacMillan, Beebe-Frankeberger, & Bocian, 2000). Because instructional decisions are being made based on data from interventions assumed to be implemented appropriately, treatment integrity data is critical in an RTI approach. Without it, there is no guarantee that the student’s response to intervention was, in fact, a product of the intervention and not a product of a failure to implement the intervention correctly. As an example of the importance of collecting treatment integrity data, Wickstrom, Jones, LaFleur, and Witt (1998) discovered that 100% of teachers in their study failed to deliver an intervention with greater than 10% treatment integrity, even after the teachers were well-trained and agreed to conduct the intervention. There are five frequently cited methods for measuring treatment integrity: direct observations, self-reports, rating scales, permanent products, and manualized treatments (Goss, Noltemeyer, & Devore, 2007).
Unfortunately, there is no agreed upon standard for how to monitor treatment integrity within an RTI approach. Ideally, treatment integrity should be monitored using multiple methods (e.g., self-report and direct observation).

*Professional development.* As a result of the variety of professionals involved in the implementation and monitoring of RTI, coupled with the variety of skills necessary to implement RTI well, continuous professional development is necessary (Brown-Chidsey & Steege, 2005). Although there has been no known empirical research examining the role of professional development in RTI implementation, several literature-based recommendations have been made. For example, Brown-Chidsey and Steege proposed literature based recommendations for developing a professional development plan that consists of three elements: (1) a schedule, (2) educator learning outcomes, and (3) indicators of mastery of RTI methods. In addition, Batsche et al. (2006) suggest that successful professional development programs adequately address the following three domains: beliefs/attitudes, knowledge, and skill.

*Examples*

Early RTI-based models were characterized primarily by the problem-solving process and data-based decision-making but did not exemplify all of the components currently thought to define RTI (e.g., absence of a formalized tier framework or systems approach). Examples of such RTI precursors include Pennsylvania’s Instructional Support Team (IST) model, established in 1990 as a means of using pre-referral intervention to reduce inappropriate special education referrals and placements (Kovaleski, Tucker, & Stevens, 1996). Another example includes Ohio’s “Intervention
Based Assessment” (IBA) model, a voluntary state-sponsored initiative introduced in 1992 (Telzrow, McNamara, & Hollinger, 2000). In the IBA model, trained Multi-Disciplinary Teams (MDT’s) at the school level utilized a problem-solving approach to address student concerns (Telzrow et al.). As part of the problem-solving approach, the following eight core components are addressed: (a) behaviorally defining the problem, (b) collecting baseline data, (c) clearly identifying the goal, (d) hypothesizing the reason for the problem, (e) creating a systematic intervention plan, (f) collecting evidence of treatment integrity, (g) collecting data on the student response to the intervention, and (h) comparing the student performance with the baseline performance (Telzrow et al.).

Although these early models established a foundation from which RTI emerged, later models began to incorporate the features that currently conceptualize RTI (e.g., tiered models). For example, the Heartland Area Educational Agency 11 (also known as Heartland) model used a four-tier problem-solving model to address student concerns (Fuchs et al., 2003). At each tier, practitioners determine the level of the problem, analyze its potential causes, develop an intervention, monitor student progress, evaluate intervention effectiveness, and plot future actions (Fuchs et al., 2003).

Minneapolis Public Schools also has refined an RTI model beginning in 1993 when they received a waiver from the Minnesota State Board of Education to explore alternate routes to special education identification (Marston, Muyskens, Lau, & Canter, 2003). The Minneapolis RTI process is based on a sequence of problem-solving steps which includes specifically defining a student’s problem, generating and implementing intervention strategies, monitoring intervention effectiveness and evaluating outcomes,
and continuing the problems-solving cycle as necessary (Marston et al.). This problem-solving process occurs on three different tiers as needed: classroom level, problem-solving team level, and special education referral level (Marston et al.). This model was phased into the district gradually in 1994, with implementation schools receiving ongoing professional development and training (Marston et al.).

The Illinois Flexible Service Delivery System Model (FSDS) is another multi-tiered model of service delivery to meet the needs of all students (Peterson, Prasse, Shinn, & Swerdlik, 2007). The basic principles of FSDS include prevention and early intervention to address academic and behavioral needs, parent involvement, systematic problem-solving process, data-based decision-making, research-based interventions, professional collaboration, progress monitoring, and special eligibility determination based on a student’s RTI (Peterson et al.). Upon initial approval of this model by the Illinois State Board of Education in 1995, six districts were approved as FSDS sites, a figure that rose to 90 after 10 years (Peterson et al.). Peterson et al. describe four “sustaining structural elements” that have facilitated success of this model over the years: (1) building-based commitment and involvement, (2) communication and learning networks, (3) professional development, and (4) an evaluation plan.

Idaho’s statewide implementation of RTI is called the Results Based Model (RBM). This model began in 1997 in three elementary schools when the state department of education was awarded a State Improvement Grant (Callender, 2007). As of 2005, approximately 150 schools were participating (Callender). The model integrates a problem-solving and standard protocol approach in four levels: (a) general education
for all students, (b) standard protocol small group interventions for intensive and strategic students, (c) problem-solving and individualized interventions available to students as needed, and (d) special education and intensive long-term services (Callender). The model is comprised of 8 big ideas or key practices: addressing the system, problem-solving teams, parental involvement, functional assessment, outcome-oriented intervention, ongoing progress monitoring, systematic data-based decision-making, and dual discrepancy (Callender).

Schoolwide Positive Behavior Support

Conceptual Framework

SWPBS is a comprehensive approach designed to promote the appropriate behaviors of all students and enhance the capacity of schools and families to design positive environments for students (OSEP Center on Positive Behavior Interventions and Supports, 2004). SWPBS has been defined as, “…the application of positive behavioral interventions and systems to achieve socially important behavior change” (Sugai, Horner, et al., 2000, p.133)

Components

As previously mentioned, SWPBS shares many of the same conceptual framework and components of RTI. For example, SWPBS is based on a three-tiered model of prevention and intervention, although the tiers are sometimes referred to as Primary, Secondary, and Tertiary as opposed to Tier I, Tier II, and Tier III (see the right side of the cone in Figure 1). In addition, SWPBS relies on the same foundation of effective professional development, evidence-based interventions, and treatment
integrity. Despite the similarities, SWPBS has several unique features necessary for successful implementation. Specifically, Lewis and Sugai (1999) identified six essential elements of a SWPBS model that are necessary at a systemic level.

Statement of purpose. A statement of purpose serves as a guide for how a school makes decisions and conducts business. Every school should develop a statement of purpose that encompasses its function and summarizes its approach to teaching and learning (Lewis & Sugai, 1999). The statement should be concise, positively worded, inclusive, and address both academic and behavioral outcomes (Lewis & Sugai). For example, Lewis and Sugai provide the following example of a quality statement of purpose: “At K&R School, students and staff 1) place highest value on academic and social success; 2) strive for proactive and safe learning and teaching environments; 3) foster partnerships with students, families, and communities; and, 4) emphasize what works.”

Schoolwide behavioral expectations. Every school should generate a unique list of behavioral expectations. That is, the SWPBS team or school staff should identify three to five behavior expectations relevant to their specific schools (Lewis & Sugai, 1999; McKevitt & Braaksma, 2008). These expectations should be positively worded, brief, memorable, and developmentally appropriate (McKevitt & Braaksma). Furthermore, the expectations should be defined as specific, observable behaviors for each location in the school (McKevitt & Braaksma; Lewis & Sugai). For example, “Be responsible” might mean picking up trash in the hallway whereas it might mean turning in homework in the
classroom. To be effective, the expectations should be posted throughout the school and used by all staff members.

**Procedures for teaching schoolwide expectations.** Once the behavioral expectations are clearly defined, they should be taught to all students using direct and explicit instruction that (a) identifies the expectations being taught, (b) models the expectations as well as non-examples, and (c) allows students the opportunity to practice the expected behaviors repeatedly in an instructional setting with corrective feedback and praise (McKevitt & Braaksma, 2008). Each school should have a plan to provide initial instruction for all students at the beginning of the school year followed up by booster sessions based on need (McKevitt & Braaksma). To increase the likelihood of students generalizing the expectations to other contexts, teaching procedures should include a variety of examples and practice in diverse environments.

**Procedures for encouraging expected behaviors.** It is important not only to teach the expectations, but also to provide incentives to encourage students to demonstrate them in the natural setting (Lewis & Sugai, 1999). For example, schools can establish an incentive system that provides tangible or non-tangible rewards for exhibiting expected behaviors on the student, classroom, grade-level, or whole school level. These incentive systems should be easy and efficient for staff members to manage. It may be useful to include students in the process of determining the incentives, in order to guarantee they will be reinforcing (McKevitt & Braaksma, 2008). Over time, the incentive systems
should be faded to include more internal, infrequent, and unpredictable reinforcement (Lewis & Sugai).

*Procedures for discouraging problem behavior.* Even when students are explicitly taught expectations and provided reinforcement, the expectations will occasionally be broken (Lewis & Sugai, 1999). In these cases, it is important to develop a consistently implemented plan with clearly defined examples of problem behaviors and specific guidelines for determining which consequences should be provided for which behaviors (Lewis & Sugai). It is important that the incorporation of consequences also includes a teaching component and is easy for staff to use consistently and efficiently (McKevitt & Braaksma, 2008).

*Procedures for record keeping and decision-making.* It is important for schools to collect data in order to determine the effectiveness of all tiers of support (Lewis & Sugai, 1999). These data should reflect both student-level outcomes and systems-level implementation. Tracking office discipline referrals (ODRs) is an efficient, effective way to monitor student-level outcomes. ODRs, used as an index of both student behavior and disciplinary systems within a school, represent events in which (a) a student engaged in a behavior that violated a school rule or norm, (b) this violation was identified by a school staff member, and (c) a written product was created to record the event and a consequence was delivered by an administrator (Sugai, Sprague, Horner, & Walker, 2000). These data should be reviewed regularly as a formative tool (i.e., to look at the most common types, locations, and times of problem behaviors to guide intervention efforts) as well as a summative tool (i.e., to determine the impact of SWPBS over the
year) (McKevitt & Braaksma, 2008). These data can also be used to determine who needs Tier II or Tier III support. Typically, if a student has 3-5 ODR’s in a year but isn't displaying significantly dangerous behaviors, Tier II intervention may be warranted. In contrast, if a student has six or more ODR’s in a year or who demonstrate significantly dangerous or disruptive behaviors, Tier III support may be needed (McKevitt & Braaksma).

Irvin, Tobin, Sprague, Sugai, and Vincent (2004) explored the validity of ODRs and determined sufficient evidence to justify their use in assessing school-wide behavioral climate, the effectiveness of behavioral intervention programs, and differing needs across schools in developing positive behavioral climates. However, concerns have been raised regarding the validity of using ODR data to make decisions regarding individual students as well as variations in ODRs based on the unique manner in which a school defines, administers, and documents ODRs (Irvin et al.). Therefore, caution must be exercised in exclusively relying on ODRs for decision-making.

In addition to collecting student-level data, systems-level implementation data also should be collected so schools can identify areas in which they need to improve. One tool for doing so is the SchoolWide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001; available online at www.pbis.org) This tool, which is designed to assess overall SWPBS implementation, involves an independent observer conducting records reviews, school observations, and interviews with students, staff, and administrators. It yields a summary score for each of the seven areas and an overall mean score. The goal is to obtain an 80% or higher on the instrument.
Examples

The Florida Positive Behavior Support Project (FLPBS) trained 81 schools in SWPBS in 2002 and 2003, 85 schools in 2004, and over 100 schools in 2005 (Kincaid, Blasé, Childs, & Wallace, 2007). The Florida model includes four tiers (Florida Department of Education, 2002). First, school-wide or universal supports are aimed towards all students and staff and are designed to meet the needs of 80% of the school population. Second, classroom supports represent school-wide expectations and pre-planned strategies used within classrooms. Third, targeted group supports address the behavioral issues of groups of students. Finally, individual student supports reflect school-wide supports in conjunction with team-based strategies for addressing the behavior problems of individual students. These supports may include functional behavior assessments and individual interventions. Overall, a representative school-based team – which meets at least monthly – is responsible for developing, implementing, and monitoring the school’s SWPBS plan (Florida Department of Education). SWPBS is viewed as a proactive, collaborative, educative, functional process (Florida Department of Education).

Maryland also has a SWPBS model, developed by a statewide collaboration in 1998 (Barrett et al., 2008). By late 2006, Maryland was implementing the SWPBS model in 33% of its schools and 100% of its 24 districts (Barrett et al.). The Maryland SWPBS model includes essential features of SWPBS such as a tiered model of support, data-based decision making, proactive school discipline, and implementation fidelity measures. It is coordinated on the state, district, and school levels. At the state level, a
leadership team coordinates and supports the implementation, training, and sustainability of SWPBS and an advisory group works to gain political support (Barrett et al.). At the local level, those districts that have a “critical mass” of buildings participating in SWPBS (5 of the 24 districts) have established teams led by a SWPBS coordinator to provide support and planning (Barrett et al.). Finally, at the school level, a building team—consisting of four to five teachers, an administrator, a team leader (e.g., school psychologist), and an external coach—develops local data collection, intervention, and decision-making processes (Barrett et al.).

A final exemplar of a state-wide SWPBS model comes from North Carolina. Since the establishment of a Positive Behavior Support Center in the fall of 2000, the number of schools implementing SWPBS in North Carolina increased to 548 during the 2006-2007 school year (Irwin & Algozzine, 2008). The North Carolina SWPBS initiative is a team-based system that incorporates a common approach to discipline that is proactive, instructional, and outcome-based. A state-level Positive Behavior Support Leadership Team has a focused Action Plan and meets regularly to refine the work being done in North Carolina’s SWPBS schools (Irwin & Algozzine). In addition, three SWPBS Regional Coordinators co-lead three subcommittees that are working on coaching and training, evaluation, and visibility and political support (Irwin & Algozzine). State funding has been available to support SWPBS initiatives in North Carolina schools.
Integrated Systems Models

Conceptual Framework & Components

ISMs involve implementing tiered models of academic (RTI) and behavioral (SWPBS) support in a simultaneous and streamlined fashion within a school, district, or state. Therefore, its components include those previously discussed for RTI and SWPBS. Interest in developing ISMs has been fueled by research demonstrating causal paths between reading and behavior outcomes for students (e.g., Kellam et al., 1998). See Figure 1 for a graphic representation of ISM.

Examples

Integrated models of academic and behavior support have been implemented on a small-scale (e.g., school, district, or regional level) in several studies. For example, Lane and Menzies (2002) studied the implementation of a Tier I ISM in an elementary school serving 298 students. They discovered significant improvements in academic outcomes (particularly for younger students) and non-significant improvements in behavioral outcomes. Given the focus of this study, however, further discussion will concentrate on large-scale (i.e., statewide) models. Two of the most documented statewide exemplars of ISM include Michigan and Ohio.

Michigan’s statewide ISM was initiated when a team of school practitioners and university faculty members formed in 1999 to develop a data-based problem solving approach to building schools’ capacity to improve academic and behavioral outcomes (Ervin, Schaugency, Goodman, McGlinchey, & Matthews, 2007). These initial contributors were interested in combining academic and behavior supports to maximize
the effect each would have. The model emphasized capacity building through (a) collaborating within building-level teams to develop problem-solving strategies; (b) establishing data systems to inform decision making; (c) establishing mechanisms to enhance sustainability; and (d) providing professional development, feedback, and incentives to support staff implementation (Ervin et al.). Site specific and locally developed action plans guided these activities (Ervin et al.). For example, building teams, external coaches, state technical assistance, and state guidance teams all played a role. Within Michigan’s model, problem-solving is deeply entrenched at both the systems-level and student-level.

The ISM of particular interest for this study is the Ohio Integrated Systems Model (OISM). Following the introduction of the statewide IBA initiative in 1992 and SWPBS initiative in 1999, an integration of these two models began in 2002 and was formalized into OISM in 2005 (Bogdan, 2006). The primary goal of OISM is, “…school improvement (with a specific focus on closing achievement gaps for all NCLB subgroups) by increasing the capacity of systems to use problem solving/data-based decision making and research-based practices to sustain high achievement over time (Graden et al., 2007, p.290),” and the secondary goals include, “…application of a problem-solving model to system level issues (as well as individual student concerns) and improvement of universal literacy and behavior instruction for all students (Graden et al., p.290).” Essentially, OISM is a prevention-based model designed to improve academic and behavioral outcomes at multiple levels (e.g., district, school, classroom, individual
student). It is characterized by a three tiered model comprised of six Key Features that are embedded within each tier. See Table 1 for a description of the six Key Features.

OISM has been supported by a variety of leadership positions including a state-level core group, three regional leads, consultants responsible for supporting district teams, district leadership teams, district coaches, and building leadership teams (for a description of the purpose and key activities of each group, see Graden et al., 2007). During the 2005-2006 school year, approximately 300 school buildings participated in OISM, supported financially through the State Improvement Grant (Graden et al.). The following year, 29 additional districts participated using grants from the Ohio Department of Education to support these activities (Graden et al.). Although the SIG funding is no longer available, many schools and districts are continuing to implement the model under the name Integrated Systems Model (ISM) and others have begun a related state-sponsored school improvement process called the Ohio Improvement Process.

Student Outcomes

Compared to the plethora of commentary on the application of RTI, SWPBS, and ISM approaches, research regarding student outcomes of field-based exemplars has been relatively scarce. Although there are no definitive reasons for this, two logical factors emerge. First, schools may be devoting their resources solely to implementation with little time or interest in evaluation. This focus on implementation may be compounded further by pressure to implement these models from legislation and policy. Second, given the systems-wide focus of these models, identifying outcomes and forming causal
Table 1

*Description of the Six Key Features of Ohio’s ISM*

<table>
<thead>
<tr>
<th>Key Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative Leadership</td>
<td>The behaviors and skills required for guiding a group to achieve goals for students. Involves providing a consistent message and communicating assumptions about educating all students.</td>
</tr>
<tr>
<td>Collaborative Strategic Planning</td>
<td>A team-based approach that uses a collaborative planning and problem-solving process to address system-level issues or student-related concerns.</td>
</tr>
<tr>
<td>Scientifically Based Research</td>
<td>Educational practices and interventions that have empirical evidence to support their effectiveness.</td>
</tr>
<tr>
<td>Data-based Decision Making</td>
<td>Frequent, reliable, and valid indicators of student performance guide the decision making at the school and student level.</td>
</tr>
<tr>
<td>Reading and Behavior Supports Across Three Tiers</td>
<td>Direct teaching, reinforcement, and correction of expected academic and social behaviors. The intensity of instruction increases with the intensity of student need. Resource allocation and coordination are essential across all systems within the district. An effective system of prevention and a continuum of supports must be in place.</td>
</tr>
<tr>
<td>Culturally Responsive Practices</td>
<td>Educational practices that are relevant, responsive, and respectful to families from diverse economic, cultural, or linguistic backgrounds.</td>
</tr>
</tbody>
</table>

attributions becomes difficult due to a host of methodological variables (e.g., lack of experimental control).

Despite these barriers, research examining student outcomes is emerging. For example, research increasingly has supported the use of RTI for a variety of purposes, including reductions in special education referrals (e.g., Burns et al., 2005; Kovaleski et al., 1996; Marston et al., 2003; VanDerHeyden, Witt, & Gilbertson, 2007), overrepresentation of minority students in special education programs (e.g., Marston et al.; Gravois & Rosenfield, 2006), overrepresentation of males in special education programs (e.g., VanDerHeyden et al.), and retentions (e.g., Kovaleski et al., 1996). More directly, the use of RTI has been associated with improved student academic outcomes (e.g., Ardoin, Witt, Connell, & Koenig, 2005; Burns et al., 2005; Peterson et al., 2007; Speece et al., 2003) and increased academic learning time (Kovaleski, Gickling, Morrow, & Swank, 1999).

SWPBS has also been demonstrated to effect student change. For example, SWPBS implementation has been documented to reduce ODRs (e.g., Barrett et al., 2008), reduce suspension rates (e.g., Barrett et al.), result in financial savings for the school (e.g., Scott & Barrett, 2004), and reduce instructional time missed due to disciplinary referrals (e.g., Scott & Barrett).

Given the focus of this study, however, the remainder of this discussion will expand on research examining the relationship between ISM and student outcomes. McGlinchey and Goodman (2008) reported on the outcomes of an integrated three-tiered model implemented in a small, high economic need, low achieving urban school district
in Michigan. Although gains in reading CBM performance were noted across all students, the results were most pronounced for kindergarteners. For example, in Spring 2002, 10% of kindergarten students met benchmark expectations on phoneme segmentation CBMs, versus 96% of kindergarten students in spring 2006. In addition, only 21% of kindergarten students met benchmark expectations on nonsense word CBMs in spring 2002 versus 90% in spring 2006. Regarding behavioral outcomes, the number of ODRs per 100 students reduced from 390 during the 2001-2002 school year to 182 during the 2005-2006 school year. During the same time period, in-school and out of school suspensions were reduced by 196 days. Finally, the percentage of 4th grade students in the district passing Michigan’s high-stakes reading assessment increased from 30% in 2000 to 61% in 2005. These figures appear promising, although causal attributions are difficult to draw due to the case study nature of the study.

McIntosh, Chard, Boland, & Horner (2006) also sought to explore the effectiveness of a small-scale ISM in improving student outcomes. Specifically, the researchers studied a school district of over 5,000 students that had been implementing an ISM for at least five years at the elementary and middle school levels. The researchers used descriptive statistics to compare to national averages the proportion of (a) third grade students in the district identified as needing additional reading support through reading CBM screening measures, and (b) kindergarten through fifth grade students in the district identified as needing additional behavioral support through ODR screening measures. Overall, they found that a smaller proportion of the students needed additional academic or behavior support (3% of third grade students and 8% of kindergarten
through fifth grade students) than national averages (40% and 14% respectively).

Although the authors suggest the results provide initial support for the use of an integrated model, they concurrently suggest future research needs to use similar comparison districts not implementing such models.

Ervin et al. (2007) studied four elementary schools (Schools A, B, C, and D) in Michigan utilizing an ISM. Among other foci, the researchers were interested in exploring whether behavioral and academic student improvements were documented during the five years of program implementation. Related to behavioral outcomes, schools B and C demonstrated low rates of ODRs across the project years. School A, however, reduced the number of ODRs per day per 100 students to the average range by 2005. School D, which began with a high number of ODRs, reduced the number but did not reach the average range. The total time taken away from instruction due to disciplinary consequences (20 minutes per ODR + time in suspensions) was reduced for Schools A and D but fluctuated for schools B and C. Schools B and C consistently reported less than 5% of students with recurrent (i.e., >6) ODRs, whereas Schools A and D documented a downward trend over the implementation years. Related to academic outcomes, Schools A and B demonstrated increases in the percentage of students meeting benchmark expectations on DIBELS measures. Although some of these results were promising, inconsistencies in improvement between schools were noted. In addition, causal attributions were difficult to draw due to the weak research design (i.e., pre-experimental repeated measures design).
Interested in similar issues, Lane & Menzies (2002) sought to explore the academic and behavioral outcomes of Tier I of a small-scale integrated model among upper- and lower-elementary school students. Participants included 298 first through sixth grade students attending an elementary school in southern California. The Tier I intervention consisted of (a) A District Literacy Plan (DLP) for providing whole class, small group, and individualized quality reading instruction, and (b) a Schoolwide Behavior Plan (SBP). Checklists of the critical components of the DLP and SBP were created and treatment integrity data were collected by the researchers via direct observation (mean treatment integrity for DLP was 90.31% and for SBP was 80.38%). Three types of outcome measures were collected: district multiple reading measures (DMR), curriculum-based assessments in reading, and an assessment of risk for antisocial behavior. Three separate repeated-measures ANOVA’s were conducted on each of these three outcome variables. Results of the study suggested significant improvements on the CBM measures across grade levels and significant improvements in the DMR measures in the lower elementary students only. However, student risk scores were not shown to evidence significant improvement over time. The authors surmised this lack of significance may have been due to more limited resources devoted to the SBP than the DLP, as well as lower treatment integrity levels. In the future, they recommended a more balanced approach to the behavioral and academic components of an integrated approach.

Finally, in a synthesis of the extant literature, Stewart et al. (2007) conducted a meta-analysis to evaluate the impact of three types of three-tiered models on academic and behavioral outcomes: academic only, behavior only, and integrated. Based on their
inclusion criteria, 11 intervention studies were included in the meta-analysis. The studies varied in number of participants, tier of intervention, grade-level, student classification (e.g., at-risk or students with disabilities) and outcome measures. Using Cohen’s (1988) effect size guidelines, researchers found small effect sizes for the reading-only ($Zr=.30$) and behavior-only ($Zr=.18$) models on reading outcomes. However, large effect sizes were found for the integrated models ($Zr=.53$) on reading outcomes. In addition, a moderate effect size was found for the integrated model on behavioral outcomes ($Zr=.31$), whereas a small effect size was found for the behavior model on behavior outcomes ($Zr=.28$). There was no effect size for reading models on behavior outcomes. Overall, results of this meta-analysis suggest an integrated model is more effective than either an academic or behavioral model in isolation at improving student academic and behavioral outcomes. The authors propose that addressing problem behaviors may increase the likelihood of responding appropriately to reading instruction.

Implementation and Student Outcomes

Implementation Research

Research has suggested that even when supported by legislation or policy, most change efforts in education over the past twenty-five years have limited implementation success (e.g., Berends et al., 2002). In fact, Mann (1978) found a success rate of only 20% for planned changes in educational programs and current data continue to support this trend (e.g., McDermott, 2000). Although partially determined by the complexity of the change initiative, it has been estimated that full implementation may take anywhere from 3 to 10 years to achieve (Fullan, 2007).
Although the definition may seem intuitive, it is important to specify exactly what is meant by the term “implementation.” Fixsen, Naoom, Blasé, Friedman, and Wallace (2005) define implementation as “a specified set of activities designed to put into practice an activity or program of known dimensions.” Implementation integrity (aka implementation fidelity) is a term that refers to the degree to which specific procedures of a program are implemented as intended (Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993). For the purposes of this study, implementation integrity is distinct from treatment integrity. Whereas the former will refer to the integrity with which a comprehensive program (e.g., RTI, SWPBS, and ISM) is implemented within a system, the latter will refer to the integrity with which a single intervention is implemented for a specific student. Given the comprehensive nature of RTI, SWPBS, and ISM, the focus of this discussion will center on implementation integrity.

Dane & Schneider (1998) examined the implementation integrity of prevention programs published between 1980 and 1994 that met their inclusion criteria. They coded the 162 studies in their final sample, and obtained inter-rater reliability of .84 on a sample of 15% of the studies. They found that only 24% of the studies provided documentation of implementation integrity. This rate is fairly consistent with other estimates. For example, Dane and Schneider noted that their findings were comparable, “…to the 20% found by Peterson, et al. (1982) in 539 experimental studies published from 1968 to 1980 in the Journal of Applied Behavior Analysis, the 18.1% found by Moncher and Prinz (1991) in 359 treatment outcome studies published in clinical psychology, psychiatry, behavior therapy, the 6% found by Rogers-Weise in 88 group-design parent training

Based on these figures, it is easy to assume that implementation levels of change programs have been less than desirable, and there is no reason to think that implementation of RTI, SWPBS, or ISM will reverse this trend. In fact, related to the Illinois RTI model, Peterson et al. (2007) noted that ensuring the model is implemented with integrity is the single most difficult issue they face. Obviously, implementing the previously mentioned components of RTI, SWPBS, and ISM in a school will likely require significant individual and systemic change. Although some schools have documented high levels of implementation integrity (e.g., McIntosh et al., 2006), others are finding it difficult (e.g., Peterson et al.). Although schools may collect treatment integrity data to verify the implementation of specific interventions at the student level – and this is of critical importance – there is little monitoring of implementation at the systems level. This is concerning because when implementation integrity data is not collected, non-significant results may be interpreted as a failure of the program when in fact it may be due to a failure to fully implement the program. (Dane and Schneider, 1998). In addition, any positive effects are likely to be underestimated if the program is not implemented as intended. Given these concerns, coupled with the lack of research examining the relationship between RTI/SWPBS/ISM and outcomes, it is relevant to turn to research on other comprehensive school reform (CSR) initiatives to explore the relationship between implementation integrity and outcomes.
Relationship between Implementation Level and Outcomes in CSR

It is clear that implementation integrity is not frequently assessed. But, when it is, how does it relate to student outcomes? Dane and Schneider (1998) found that in several studies, significant program effects were found only when the programs were implemented as planned. Other studies in the psychological and prevention literature substantiate that reduced implementation integrity is associated with reduced improvement in outcomes (e.g., Botvin, Baker, Dusenberry, Tortu, & Botvin, 1990). Given the paucity of research directly examining the relationship between RTI, SWPBS, or ISM implementation integrity and student outcomes, it is important to examine this issue in the context of broader CSR initiatives.

The Urban Institute (2007) conducted a study which addressed (a) the effectiveness of a CSR at improving student outcomes, and (b) the relationship between implementation of a CSR and student outcomes. In 2001, the Baltimore City Public School System unveiled a plan to create eight “innovation high schools” and nine “community high schools” based on a CSR model that addressed three guiding principles: strong academic rigor; small supportive structures; and effective, accountable instruction and leadership. For the study, a sample of six innovation schools, 12 neighborhood schools, and five traditional schools during the 2005-2006 school year was used. The researchers created two measures to assess implementation of the guiding principles of the CSR, one that assessed support for students and one that assessed the teaching and learning environment. Using multiple regression analyses, the researchers found that students in schools with higher implementation of student support scored
significantly higher on High School Assessment (HSA) English and Algebra measures and also had higher attendance rates. In addition, students in schools implementing higher levels of positive teaching and learning environments demonstrated higher HSA Algebra scores.

Datnow et al. (2000) also examined the relationship between CSR implementation and student outcomes. Specifically, they studied the implementation of Core Knowledge Sequence, a CSR that specifies a planned progression of topics to teach across all subjects during the elementary years, in four schools. Using classroom observer ratings, the researchers found that the CSR was implemented successfully in 3 of the 4 schools. Specifically, three schools exhibited an average of 70% implementation, whereas the fourth school exhibited 27% implementation. The researchers discovered that the three high implementing schools evidenced improvements in student achievement compared to students from control schools, whereas the low implementing school did not. The researchers concluded that this difference is likely a result of implementation differences.

Nunnery (1997) also found that level of implementation was positively related to student achievement. Specifically, the researcher designed a quasi-experimental study aimed at exploring implementation of various configurations of a CSR designed to assist at-risk students. Results suggested that schools implementing all four critical components of the program produced positive student achievement effects, whereas schools that failed to make the organizational changes to support two components of the program performed worse than comparison schools. The researcher concluded that in
order to effect positive change, schools need to commit to implementing programs in their entirety and supporting them in the school infrastructure.

Despite the general consensus on the association between implementation integrity and student outcomes, there also have been results to the contrary. For example, Nunnery, Ross, & McDonald (2006) conducted a randomized field study to examine the impact of implementation of a comprehensive reading program on student achievement. Overall, they found that the program did produce achievement gains for students despite suboptimal implementation integrity (i.e., classrooms using the program performed better than control classrooms). However, quality of implementation integrity did not predict student achievement, except in students with learning disabilities.

*Relationship Between Implementation Integrity and Outcomes in RTI, SWPBS, and ISM*

Gansle & Noell (2007), in a discussion of treatment integrity, remarked, “…the research between [treatment integrity] and intervention outcomes is neither sufficiently systematic nor programmatic to allow for a synthesis of research that provides a positive link between them (p.247).” However, they continue to note that, “It appears that as [treatment integrity] begins to break down, the risk that treatment will deteriorate increases, and that the degradation of different components may affect outcomes differently for different students (p.247).” Gansle and Noell conclude that because schools do not have the time or resources to assess implementation of every step of every intervention process, it would be useful to identify and assess those components that have been demonstrated by research to be critical to outcomes. However, since such analyses
are typically unavailable for most interventions used in schools, it would be practical to identify those steps thought to be essential for success and assess them.

Along similar lines, Kovaleski (2007) identified two critical lines of RTI research needed. One, which is related intricately to the purpose of this study, is examining, “…the extent to which the entire comprehensive [RTI] model produces positive outcomes when all individual components are implemented with effectiveness and when the process is operated with an appropriate level of efficiency (p.640).” When doing so, Kovaleski recommends giving attention to the dependent variable used. Although a decrease in the number of students identified as needing special education services has often been used as an outcome measure, researchers have questioned the validity of this variable for assessing outcomes (e.g., Vaughn & Fuchs, 2006). Given the pressure schools face in complying with NCLB, Kovaleski recommends research using outcomes linked more closely to statewide assessments or other variables with practical significance.

Based on the comments by Gansle and Noell (2007) and Kovaleski (2007), it is clear that more research linking implementation integrity levels to student outcomes and identifying key implementation components related to improved outcomes in warranted for RTI, SWPBS, and ISM. Unfortunately, this undertaking has not been fully accomplished for ISMs. Only one known study attempts to explore this issue on Intervention Based Assessment (IBA), a precursor of RTI that shares some of its features. Specifically, Telzrow et al. (2000) conducted a study on Ohio’s IBA model aimed at providing an objective analysis of the relationship between IBA quality and student
outcomes. Participants in this study included 227 of 329 multidisciplinary teams (MDT’s) in the state of Ohio who participated in the statewide IBA initiative during the 1996-1997 school year. As part of their participation in the IBA initiative, Multi-Disciplinary Teams (MDTs) submitted a Problem Solving Worksheet and Evaluation Team Report documenting the problem-solving on one student. The Problem Solving Worksheet asked MDT’s to record information pertaining to each of the 8 core IBA components and the Evaluation Team Report involved describing and analyzing the student’s concerns as well as describing interventions and how they were monitored. MDT’s were asked to submit “best case” documentation that would reflect their best quality implementation of the 8 core IBA components. Telzrow et al. evaluated each MDT’s documentation using a Likert scale and scoring rubric to assess the 8 core components and student outcomes, and inter-rater agreement data showed high levels of consistency using the scoring rubric.

The researchers found that ratings of students’ level of progress on their goals indicated positive student change as a result of the IBA process. Results also indicated that the core components with the highest mean scores were “behavioral definition of the problem” and “clearly identified goal”, while “hypothesized reason for the problem” and “treatment integrity” were the components with the lowest mean scores. In addition, a statistically significant positive relationship was discovered between student outcomes and mean fidelity ratings for six of the eight core IBS components. A stepwise multiple regression analysis also indicated that a combination of two core components, “clearly identified goal” and “data indicating student response to intervention” were significant
predictors of student outcomes, accounting for 8% of the variance in outcomes. Finally, there was no relationship found between the number of years the MDT had utilized the IBA process and the degree of implementation fidelity.

Although this study provided useful information regarding the implementation of a state-initiated outcomes-based intervention model, there were several limitations present. First, it is possible that those MDT’s that submitted documentation may have been more proficient or committed to the process than other MDT’s, resulting in sample bias. In addition, due to the method of data collection (written case documentation submitted by MDT’s) it is possible that some teams accurately implemented more of the core components but did not record the information on the written reports. Also, given the shift from IBS to OISM, the implementation integrity measures need to be broadened to include the six Key Features of OISM. Finally, given the focus on systems-level change that defines OISM, it is important to assess the effect of systems-level implementation (rather than case or student-level implementation) on student outcomes. It is the intent of this study to clearly address the latter three limitations of the Telzrow et al. study, thereby improving upon the extant knowledge in the domain and setting the stage for improved practice.

Summary and Implications

Despite limited research in the domain, initial findings suggest that although academic and behavioral models of support each result in improved student outcomes, maximal benefits are obtained when they are implemented simultaneously in an ISM
(e.g., Stewart et al., 2007). However, these findings are tentative and need further replication.

In addition, implementation of RTI, SWPBS, and ISM has not been consistently or effectively monitored or linked to changes in student outcomes. Although research on other examples of CSR suggests a relationship between implementation level and improved student outcomes (e.g., Urban Institute, 2007; Datnow et al., 2000), this only has been documented for a precursor to RTI (i.e., Telzrow et al., 2000) and not yet extended to current conceptualizations of RTI, SWPBS, or ISM. It is important to better understand the degree to which components of the models are currently being implemented, as well as the degree to which student outcomes reflect differences in implementation.

Finally, given limited time and resources available for ensuring the implementation of every component of ISMs (Gansle & Noell, 2007), researchers have called for investigations of those critical components that are most highly related to improvements in student outcomes (e.g., Gansle & Noell). By identifying these components, schools can target resources to ensure they are implemented with fidelity in order to maximize the likelihood of student success.
CHAPTER III: RESEARCH METHODS

Overview

This chapter provides an outline of the methods used to conduct the study. Specific descriptions of the participants, instruments, dependent variables, and data analysis procedures are outlined below.

Participants

Participants included all students in grades one through six who were given Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (ORF) spring benchmark assessments during the 2006-2007 and 2007-2008 school years and attended schools implementing Ohio’s Integrated Systems Model (OISM) that met predetermined inclusion criteria. Specifically, the participating schools must have done all of the following during the 2007-2008 school year: (a) implemented OISM, (b) received consultative support for OISM through their local State Support Team (SST), (c) had the Implementation Evaluation Tool (IET) administered to the school, (d) administered DIBELS ORF benchmark assessments to its students, and (e) collected schoolwide behavioral data.

Using the steps outlined in the “Procedures” section, a sample of eleven schools was obtained. However, after beginning the study, it became apparent that the IET was not administered to three of the schools during the 2007-2008 school year. These schools had been administered a previous version of the IET during the 2006-2007 school year; however, due to differences in items, results were not comparable across versions. Therefore, these schools were eliminated from further analyses.
In addition, after visually inspecting key demographic information on each school, multiple one-sample t-tests were used to obtain a more homogeneous sample. One-sample t-tests are designed to determine if the mean value of a variable is significantly different from a hypothesized value (StatSoft, Inc., 2007). In the current study, these tests were used to determine if the mean value on each of six key demographic variables across the eight remaining schools was significantly different from each school’s value on each variable. The six demographic variables of interest were selected because they were amenable to analysis (i.e., continuous data) and were deemed to be important indicators of student demographics and school environment. These variables included: average daily enrollment, percentage of economically disadvantaged students, percentage of non-white students, percentage of students with disabilities, performance index score (a score of 0-120 that reflects the achievement of every child enrolled for the academic year based on statewide assessments), and percentage of teachers with master’s degrees. A priori it was suggested that the percentage of students passing the third grade reading achievement test and the percentage of students passing the fourth grade reading achievement test also should be used as demographic variables. Upon further consideration, however, it was determined that inclusion of these variables was inappropriate and redundant because they were already incorporated into the performance index score.

Following computation of the t-tests, a chart was created to display which participating schools were significantly different from the means on each variable. Although the values on the chart are excluded to protect the identity of each school, Table
2 describes which schools varied significantly from the mean on each demographic variable. Based on examination of this chart, two schools (i.e., Schools 4 and 8 from Table 2) were eliminated from further analyses because significant differences were found on five of the variables.

As a result of these sample modifications, the final sample consisted of 2,660 student participants attending six schools. At request of participating schools, all demographic information and results are presented as means and standard deviations for each grouping of each independent variable – rather than information on individual schools – in order to maintain confidentiality. A comparison of mean demographic information for participating schools and state averages is included in Table 3.

Measures

*Implementation Evaluation Tool (IET)*

The Implementation Evaluation Tool (IET; State Improvement Grant State Steering Committee, 2007) is an instrument designed to assess implementation of the Key Features of OISM. At the school level, this tool is used to (a) determine the Key Features that are currently in place within the school, (b) identify targeted areas for improvement in collaborative strategic planning (CSP), and (c) compare implementation level across school years (IET manual, 2007).

The IET was modeled after the Schoolwide Evaluation Tool (SET-2; Sugai et al., 2001), an instrument designed to evaluate the presence or absence of critical features of
Table 2

*T-Test p-Values for the Eight Schools on Six Demographic Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Daily</td>
<td>.59</td>
<td>.88</td>
<td>.56</td>
<td>.04*</td>
<td>.16</td>
<td>.31</td>
<td>.00*</td>
<td>.03*</td>
</tr>
<tr>
<td>Enrollment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Economically</td>
<td>.21</td>
<td>.18</td>
<td>.03*</td>
<td>.00*</td>
<td>.61</td>
<td>.36</td>
<td>.10</td>
<td>.00*</td>
</tr>
<tr>
<td>Disadvantaged Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Non-White Students</td>
<td>.04*</td>
<td>.04*</td>
<td>.16</td>
<td>.01*</td>
<td>.10</td>
<td>.02*</td>
<td>.50</td>
<td>.01*</td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Students with a Disability</td>
<td>.17</td>
<td>.23</td>
<td>.04*</td>
<td>.99</td>
<td>.32</td>
<td>.79</td>
<td>.10</td>
<td>.00*</td>
</tr>
<tr>
<td>Performance Index</td>
<td>.92</td>
<td>.03*</td>
<td>.01*</td>
<td>.01*</td>
<td>.48</td>
<td>.67</td>
<td>.08</td>
<td>.00*</td>
</tr>
<tr>
<td>Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Teachers with a Master’s Degree</td>
<td>.06</td>
<td>.73</td>
<td>.04*</td>
<td>.00*</td>
<td>.32</td>
<td>.06</td>
<td>.21</td>
<td>.26</td>
</tr>
</tbody>
</table>

*indicates $p < .05$
Table 3

Comparison of Mean Demographic Information for Participating Schools versus Statewide Averages (Ohio Department of Education, 2009)

<table>
<thead>
<tr>
<th></th>
<th>School Mean</th>
<th>Statewide Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Economically Disadvantaged Students</td>
<td>23.35</td>
<td>37.7</td>
</tr>
<tr>
<td>% Non-White Students</td>
<td>10.07</td>
<td>24</td>
</tr>
<tr>
<td>% Students with a Disability</td>
<td>10.75</td>
<td>14.6</td>
</tr>
<tr>
<td>Performance Index Score</td>
<td>96.58</td>
<td>92.3</td>
</tr>
<tr>
<td>% Teachers with a Master’s Degree</td>
<td>73.80</td>
<td>n/a</td>
</tr>
<tr>
<td>Average Daily Enrollment</td>
<td>587.33</td>
<td>n/a</td>
</tr>
</tbody>
</table>

n/a indicates data were not available

School Wide Positive Behavior Support (SWPBS) and monitor implementation progress over time. An ad hoc evaluation subcommittee formed out of the State Improvement Grant State Steering Committee developed the IET in an effort to evaluate the success of ISM implementation in Ohio schools (Schaeffer, personal communication, 2008). The content and structure of the original version of the IET have been revised to better assess implementation and the instrument is now in its third edition. The most recent version of the IET has been documented to have inter-rater reliability of .94 through a process in
which two employees of an educational agency independently completed the IET on a school and compared results (Shroeder, email communication, 2009).

The IET assesses implementation of the six Key Features of Ohio’s ISM implementation (see Table 1 on page 40 for a description of each Key Feature): Administrative Leadership (AL), Collaborative Strategic Planning (CSP), Culturally Responsive Practices (CRP), Data-Based Decision Making (DBDM), Scientifically-Based Research (SBR), and Academic and Behavioral Supports Across Three Tiers (TT).

In order to evaluate these features across various levels of implementation, there are three phases of the IET (IET Manual, 2007). Phase I – Planning for Implementation – is administered to all schools in their first year of ISM implementation, as well as any schools who have not yet achieved a score of 80% on the Phase I IET (See Appendix A). Phase II – Emerging Implementation – is administered to all schools that have achieved at least an 80% on the Phase I IET (See Appendix B). Phase III – Advanced Implementation – is administered to all schools who have achieved at least an 80% on the Phase II IET (See Appendix C). Finally, Phase IV – Maintenance – does not involve any formal evaluation, and is granted to any school achieving an 80% on the Phase III IET. Each Phase of the IET builds on the Phase before. That is, all of the items on the Phase I IET are included on the Phase II IET in addition to some more advanced items. Also, all items on the Phase III IET are found on the Phase II IET in addition to some more advanced items.

The IET has been used in ISM schools in Ohio since the 2005-2006 school year. At that time, it was administered to select schools in southwest Ohio. During the 2006-
2007 school year it was piloted in several sites across the state (Schaeffer, personal communication, 2008). Finally, during the 2007-2008 school year the IET was administered to one school in each of the OISM participating districts in the southwest Ohio region, as well as every school in one of the participating districts (Schaeffer).

Administration of the IET involves collection of multiple pieces of data from the school, including: building team planning meeting schedules, notes, action plans, and data reports; school handbook; discipline code of content; written continuum of responses for behavioral violations; lesson plans; curriculum maps; DIBELS end of the year summary reports; end of the year summary reports on disciplinary infractions; professional development plans and agendas; instructional integrity checklists; and materials from a sample of individual student intervention cases (IET Manual, 2007). In addition, the individual administering the IET (usually an outside consultant or doctoral student trained on the IET) arranges a half-day visit to each school to conduct observations and semi-structured interviews with administrators, teachers, and students. Each item is ranked on a zero to two Likert-scale with specific scoring criteria provided for each item.

Outcome scores can be reported via a total score (sum of 0s, 1s, and 2s) or as a percentage of items implemented. A priori, it was anticipated that the former approach would be used for this study. However, the data were provided to the researcher using percentages. In addition, it became clear that in its actual use, IET results were provided to schools in this metric. Finally, it was discovered that when using raw scores, Phase II schools scored lower than Phase III schools because of a minimal number of possible
points, despite the high fidelity implementation on Phase II items. Because they were not
given the chance to earn points on Phase III items does not conclusively indicate that
Phase II schools were not implementing those items with integrity. Therefore, percentage
of points earned on each phase of the IET was used to assess implementation for the
study.

*Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*

DIBELS (Good & Kaminski, 2002) are standardized, individually administered,
curriculum based measures of basic reading skills designed to (a) identify children in
need of additional intervention through screening administered three to four times per
year to all students, and (b) evaluate the effectiveness of intervention strategies for select
students. Although there are multiple DIBELS measures (e.g., Initial Sound Fluency,
Letter Naming Fluency, Phoneme Segmentation Fluency, and Nonsense Word Fluency),
this study was concerned exclusively with the DIBELS Oral Reading Fluency (ORF)
measures.

DIBELS ORF measures – designed as tests of accuracy and fluency with
connected text – are typically administered in the fall, winter, and spring of the first
through sixth grade years (University of Oregon Center on Teaching and Learning,
2008). Administration of DIBELS ORF measures involves a trained administrator
having a student read aloud for one minute from three grade-level passages. The
outcome is the median number of correct words read per minute from the three passages
based on specific scoring criteria (University of Oregon Center on Teaching and
Learning).
A number of studies and research overviews have confirmed the technical adequacy of DIBELS ORF as measures of reading achievement. For example, test-retest reliability for ORF measures has been found to range from .92 to .97 and alternate form reliability has been found to range from .89 to .94 (Tindal, Marston, & Deno, 1983). In addition, criterion-related validity has been found to range from .52 to .91 (Good & Jefferson, 1998). Performance on the ORF measures also has been found to significantly correlate to performance on statewide reading achievement tests (e.g., Vander Meer, Lentz, & Stollar, 2005). Given the demonstrated technical adequacy of DIBELS ORF measures, this study used the results of these assessments as a measure of student-level reading performance.

Office Discipline Referrals

Office Discipline Referrals (ODRs) represent events in which (a) a student engaged in a behavior that violated a school rule or norm, (b) this violation was identified by a school staff member, (c) a written product was created to record the event, and (d) a consequence was delivered by an administrator (Sugai, Sprague, et al., 2000). Irvin et al. (2004) explored the validity of ODRs and determined sufficient evidence to justify the use of ODRs in assessing school-wide behavioral climate, the effectiveness of behavioral intervention programs, and differing needs across schools in developing positive behavioral climates. ODRs are often recorded in a data management system called School Wide Information System (SWIS; Educational & Community Supports, 2008). Using SWIS, the frequency of ODRs can be extracted at the individual or aggregated level and a variety of analysis techniques can be used to interpret the data. For the
purposes of this study, however, only aggregated data were available. These data were used as a measure of behavioral outcomes at the school-level.

Dependent Variables

For the first three research questions, the dependent variable represented student-level DIBELS ORF performance for the spring 2008 benchmark assessment. Raw scores – which were originally reported as the median number of correct words read per minute from three grade level reading passages – were converted to z-scores for each grade-level across schools using SPSS. A z-score is a distribution that standardizes raw scores to a mean of zero and a standard deviation of one in order to indicate the direction and degree a particular score deviates from the mean (George & Mallory, 2006). In addition, z-scores allow for different types of measurements to be comparable (Morgan, Leech, Gloekner, & Barrett, 2007). For example, in this study, transforming raw scores to z-scores allowed second grade ORF scores and sixth grade ORF scores to be comparable because both described how far a particular score was from the mean for that grade-level.

For the final research question, the dependent variable was the average number of ODRs per 100 students for each school during the 2007-2008 school year. This information was obtained through the school-entered SWIS database. Unlike the other research questions, this dependent variable represented school-level – rather than student-level – data. A priori, it was believed these data would be available for all six participating schools. However, after initiating the study it was revealed that access to the information for four of the schools was no longer available due to a change from
SWIS to another reporting system. As a result, the dependent variable was available for only two schools. See Table 4 for a summary of the variables for each research question.

Independent Variables

For the first and fourth research questions, the independent variable was overall school implementation level of ISM. Specifically, participating schools were divided into Low, Medium, and High implementers based on their standing in the bottom third, middle third, or top third of the sample, as defined by the percentage of total points earned on the IET administration during the 2007-2008 school year.

For the second research question, the implementation level for each of the six Key Features served as the independent variables. Specifically, each variable was measured using the percentage of components implemented on each of the six scales of the IET: Administrative Leadership (AL), Collaborative Strategic Planning (CSP), Culturally Responsive Practices (CRP), Data-Based Decision Making (DBDM), Scientifically-Based Research (SBR), and Academic and Behavioral Supports Across Three Tiers (3T). For a complete list of which items were determined by the scale developers to theoretically load onto which Key Features for each Phase of the IET, see Table 5. For each Key Feature, schools were divided into Low, Medium, and High implementers based on whether their implementation percentage placed them in the bottom third, middle third, or top third compared to other participating schools on that scale.

For the third research question, the two independent variables were school implementation level of the (a) academic and (b) behavioral components of ISM. On each phase of the IET, items that theoretically loaded onto the (a) academic, and
Table 4

*Summary of Variables by Research Question*

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Dependent Variable</th>
<th>Independent Variable(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Are there significant differences in student academic outcomes based on school ISM implementation level?</td>
<td>Student-level DIBELS ORF scores for the spring 2008 benchmark assessment, converted to z-scores</td>
<td>Overall school implementation level (i.e., high, medium, or low) of ISM</td>
</tr>
<tr>
<td>2) Are there significant differences in student academic outcomes based on school implementation level of each of the Six Key Features of Ohio’s ISM?</td>
<td>Student-level DIBELS ORF scores for the spring 2008 benchmark assessment, converted to z-scores</td>
<td>School implementation level (i.e., high, medium, or low) for each of the six Key Features</td>
</tr>
<tr>
<td>3) Are there significant differences in student academic outcomes based on school implementation level of the (a) academic components of ISM and (b) behavioral components of ISM?</td>
<td>Student-level DIBELS ORF scores for the spring 2007-2008 benchmark assessment, converted to z-scores</td>
<td>School implementation level (i.e., high, medium, or low) for (a) the academic components, and (b) the behavioral components of ISM</td>
</tr>
<tr>
<td>4) Are there significant differences in school behavioral outcomes based on school ISM implementation level?</td>
<td>The average number of ODRs per 100 students for each school during the 2007-2008 school year</td>
<td>Overall school implementation level of ISM (high, medium, or low)</td>
</tr>
</tbody>
</table>
Table 5

*Theoretical Loadings for the Six Scales of the IET*

<table>
<thead>
<tr>
<th>Scale</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>1, 3, 6</td>
<td>1, 3, 4, 5, 9, 10, 11, 12</td>
<td>1, 3, 4, 5, 9, 10, 11, 12</td>
</tr>
<tr>
<td>CSP</td>
<td>1, 2, 4</td>
<td>1, 2, 3, 6, 9, 10</td>
<td>1, 2, 3, 6, 9, 10, 33</td>
</tr>
<tr>
<td>CRP</td>
<td>5, 6</td>
<td>7, 8, 9, 12</td>
<td>7, 8, 9, 12,</td>
</tr>
<tr>
<td>DBDM</td>
<td>5, 7, 8</td>
<td>7, 14, 15</td>
<td>7, 16, 17, 18, 19, 34</td>
</tr>
<tr>
<td>SBR</td>
<td>6</td>
<td>11, 12, 13, 24, 26</td>
<td>11, 12, 13, 14, 15, 28, 30</td>
</tr>
<tr>
<td>TT</td>
<td>6, 9</td>
<td>11, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,</td>
<td>23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33</td>
</tr>
</tbody>
</table>

(b) behavioral components of ISM were first identified. Specifically, items including the terms “academic,” “reading,” or “literacy,” in the absence of any of the behavioral terms were identified as the academic items. In contrast, items including the terms, “behavioral,” “expectations,” “consequences,” and “behavior,” in the absence of any of the academic terms were identified as the behavioral items. See Table 6 for a complete list of items representing the academic and behavioral components for each Phase of the
IET. Schools were identified as Low, Medium, or High implementers on each component based on whether the percentage of points earned for that component was in the bottom third, middle third, or top third compared to other schools in the sample.

Table 6

Theoretical Loadings for the Six Scales of the IET on Academic and Behavior Components

<table>
<thead>
<tr>
<th>Scale</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic</td>
<td>7</td>
<td>14, 24, 25, 26, 27</td>
<td>16, 18, 28, 29, 30, 31</td>
</tr>
<tr>
<td>Behavior</td>
<td>8, 9</td>
<td>15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 32</td>
<td>17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 32</td>
</tr>
</tbody>
</table>

Design

A causal-comparative research design was used to answer all research questions. This type of ex post facto research design is used to compare two or more existing (i.e., non-manipulated) groups on one or more dependent variables (Wiersma, 2000).

Unlike experimental and quasi-experimental research, causal comparative research does not involve manipulation of the independent variable. Unlike correlational studies, causal comparative studies involve comparing two or more groups rather than looking for relationships between continuous variables.
Procedures

This study relied on existing data sets previously collected by Ohio’s State Improvement Grant (SIG) coordinator. Upon receiving a brief proposal, a designee of the SIG coordinator contacted the building representative from each school that met the predetermined inclusion criteria to determine if their contact information could be released to the researcher. The designee of the SIG coordinator also sent a reminder one week after the initial email. The researcher contacted those building representatives who provided permission to be contacted. Specifically, a phone call was made and a copy of the consent letter (see Appendix D) and a return envelope were sent through the mail. Participating schools were asked to provide written consent for the release of two sets of data: (a) school-level data to be released by the SIG coordinator (or a designee) to the researcher (i.e., IET data, aggregated DIBELS data and SWIS data), and (b) student-level DIBELS data with names removed to be released by the building representative via researcher-provided instructions.

Upon receiving the released data from these sources and obtaining written IRB approval, the researcher entered the data into SPSS 14.0 software for analysis. Data were entered by the primary research and rechecked for accuracy by an undergraduate psychology research assistant familiar with using SPSS and trained in entering data.

As previously mentioned, schools were divided into Low, Medium, and High implementers based on whether their implementation percentage placed them in the top third, middle third, or bottom third of the sample for each independent variable. For some variables (e.g., IET, CRP) division of participants into groups using this method
proved straightforward. In these cases, no two schools had the same percentage of implementation and thus students attending two schools were assigned to each of the three levels of the independent variable. In other cases (e.g., SBR, DBDM), however, assignment to groups became more difficult because several schools scored the same percentage on the independent variable and two schools could not be equitably assigned to each grouping of the independent variable.

In these circumstances, the rules used to assign participants to groups were: (a) participants attending schools that scored the same value on the independent variable would be in the same group, and (b) if there were two options for group placement, the higher implementing option would always be selected to maintain consistency. For example, for the variable AL the six schools had the following values: 50, 50, 44, 38, 38, and 25. In this circumstance, the students attending the schools that scored 50 would clearly be in the High implementing group and the students attending the school that scored 25 would clearly be within the Low implementing group. Based on the decision rules, the students attending the schools that scored a 38 were placed in the Medium implementing group and students attending the school that scored a 44 were placed in the High implementing group.

Data Analysis

Descriptive Statistics

Descriptive statistics were computed to gather more information about the characteristics of the variables. First, means and standard deviations for each of the independent variables were computed. Next, means and standard deviations for the
dependent variable across all levels of each independent variable were computed. In order to protect the identity of participating school districts, descriptive statistics are presented as means for each level of the independent variable (rather than information on individual schools).

**Analysis of Covariance**

A priori, the analysis deemed most appropriate for answering the research questions was univariate Analysis of Covariance (ANCOVA). This technique is used to compare more than two groups simultaneously on a dependent variable while controlling for a covariate (Stevens, 1999). Covariates are variables that are significantly correlated with the dependent variable but have low correlations with themselves (Stevens). For the first three research questions, it was anticipated that the z-score for each student’s DIBELS ORF score in spring 2007 would be used as a covariate. For the final research question, it was anticipated that the average number of ODRs per 100 students for each school during the 2006-2007 school year would be used as a covariate.

However, as the study progressed, compelling reasons emerged suggesting this analysis technique was inappropriate given the unique data set. Specifically, one of the prerequisite assumptions for running this analysis was not tenable. The assumption of homogeneity of regression slopes is considered so important that violations are severe (e.g., Maxwell, Delaney, & O’Callaghan, 1993; Rogosa, 1980). This is because violations suggest different effects of the independent variable on the dependent variable for different levels of the independent variable. To test for the assumption of homogeneity of regression slopes, the interaction effects between the independent
variables and covariate were examined. Whereas non-significant $F$-values indicate the assumption is tenable (Stevens, 1999), significant $F$-values ($p < .001$) were found across all independent variables. Violation of this assumption was confirmed visually by examining scatterplots, which revealed non-parallel (and intersecting) slopes.

**Factorial Analysis of Variance**

Because homogeneity of regression slopes is viewed as a critical assumption that must be met to use ANCOVA (e.g., Maxwell et al., 1993; Rogosa, 1980), the violation of this assumption was deemed to preclude use of this technique. As such, alternative tools for analyzing the data were sought. After reviewing the literature, it was decided that the best option was to perform multiple two-way factorial ANOVAs using as independent variables both the independent variable of interest and a blocked variable representing the covariate in each analysis (Lomax, 2000).

Factorial ANOVA examines the effect of two or more independent variables on a single dependent variable (Stevens, 1999). In addition to providing the capability to explore the main effects of each independent variable on the dependent variable, the interactive effects of the independent variables on the dependent variable are of interest (Stevens). An interaction is revealed when the effect of an independent variable on the dependent variable is not the same for all levels of another independent variable (Stevens). In the current study, the particular interest is the interaction between school implementation level (i.e., for the overall IET, each of the six Key Features, and the academic and behavior components) and spring 2007 student DIBELS ORF performance. This method revealed whether differences between groups in 2008 student reading
performance based on school implementation level differ by 2007 student reading performance group.

Statistical Assumptions

There are three statistical assumptions of factorial ANOVA. First, the observations must be normally distributed on the dependent variable within each group. A Kolmogorov-Smirnoff test was run on the dependent variable for each grouping of each independent variable (Rutheford, 2001). See Table 7 for a summary of tenable and non-tenable results for each analysis. Because the F-statistic is robust to violations of the normality assumption (Stevens, 1999), non-normality did not prevent analysis and interpretation.

Second, the population variances for the groups must be equal. In order to test this assumption, the Levene test was used. This test was selected because it is the most robust against non-normality (Stevens, 1999). All results suggested assumption violations (see Table 8). Interestingly, tests of homogeneity of variance are able to detect assumption violations more powerfully in larger samples, but violations are less likely to result in error than in small samples (Warner, 2008). Therefore, the violations detected tend to be the ones that matter the least. As a result, researchers have suggested additional procedures for determining if correction for violation of this assumption is needed. For example, Keppel, Saufley, & Tokunaga (1992) suggest determining the degree to which variances in the conditions are different using an index, $F_{\text{max}}$, which involves dividing the largest variance in the experiment by the smallest variance in the
<table>
<thead>
<tr>
<th>Implementation Level</th>
<th>K-S</th>
<th>Df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.02</td>
<td>644</td>
<td>.20</td>
</tr>
<tr>
<td>Medium</td>
<td>.04</td>
<td>894</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>1122</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td></td>
<td>AL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.04</td>
<td>442</td>
<td>.20</td>
</tr>
<tr>
<td>Medium</td>
<td>.02</td>
<td>644</td>
<td>.20</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>1574</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td></td>
<td>CRP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.02</td>
<td>644</td>
<td>.20</td>
</tr>
<tr>
<td>Medium</td>
<td>.04</td>
<td>1124</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>892</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td></td>
<td>CSP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.02</td>
<td>803</td>
<td>.20</td>
</tr>
<tr>
<td>Medium</td>
<td>.03</td>
<td>1185</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>672</td>
<td>&lt;.01</td>
</tr>
<tr>
<td></td>
<td>DBDM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.04</td>
<td>361</td>
<td>.19</td>
</tr>
<tr>
<td>Medium</td>
<td>.03</td>
<td>1177</td>
<td>.01*</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>1122</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td></td>
<td>SBR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.02</td>
<td>1096</td>
<td>.20</td>
</tr>
<tr>
<td>Medium</td>
<td>.04</td>
<td>442</td>
<td>.20</td>
</tr>
<tr>
<td>High</td>
<td>.04</td>
<td>1122</td>
<td>&lt;.01*</td>
</tr>
</tbody>
</table>
Table 7 (continued)

*Kolmogorov-Smirnoff Results by Group*

<table>
<thead>
<tr>
<th></th>
<th>TT</th>
<th>ACADEM</th>
<th>BEHAV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
</tr>
<tr>
<td>Low</td>
<td>644</td>
<td>452</td>
<td>644</td>
</tr>
<tr>
<td>Medium</td>
<td>.04</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>High</td>
<td>.20</td>
<td>.20</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>894</td>
<td>1086</td>
<td>672</td>
</tr>
<tr>
<td></td>
<td>&lt;.01*</td>
<td>.02*</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td></td>
<td>1122</td>
<td>1122</td>
<td>1344</td>
</tr>
</tbody>
</table>

*indicates $p < .05$

experiment. If $F_{max}$ exceeds 3.0, then the researcher should use 0.025 (rather than 0.05) as the critical value for $F$ to correct for the assumption violation (Keppel et al.). In the current project, $F_{max}$ was calculated to be 1.90. Because this is less than the cutoff of 3.0, continued analysis using a predetermined significance value of $p = 0.05$ was deemed appropriate.

Finally, the observations must be independent across groups. A priori, it was determined that intraclass correlations would be computed to assess the assumption of independence. After further reflection, however, it was determined that the independence
Table 8

Levene Test Results for All Analyses

<table>
<thead>
<tr>
<th>Analysis</th>
<th>$F$</th>
<th>df1</th>
<th>df2</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IET</td>
<td>5.13</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>AL</td>
<td>15.11</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>CRP</td>
<td>4.25</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>CSP</td>
<td>5.02</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>DBDM</td>
<td>2.78</td>
<td>8</td>
<td>2651</td>
<td>.01*</td>
</tr>
<tr>
<td>SBR</td>
<td>15.45</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>TT</td>
<td>5.13</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>ACADEM</td>
<td>2.15</td>
<td>8</td>
<td>2651</td>
<td>.03*</td>
</tr>
<tr>
<td>BEHAV</td>
<td>3.70</td>
<td>8</td>
<td>2651</td>
<td>&lt;.01*</td>
</tr>
</tbody>
</table>

*indicates $p < .05$

Assumption can be assumed because (a) each participant attended one, and only one, participating school during the time frame of interest, and (b) implementation of ISM and student performance at one school should not be affected by implementation of ISM or student performance in other schools.

**Measurement Assumptions**

The two measurement assumptions necessary to compute a factorial ANOVA were deemed to be present in the study. First, the independent variables were categorical variables that were considered nominal for the purposes of the analysis. Second, the
dependent variables were continuous variables measured on an interval scale of measurement.
CHAPTER IV: RESULTS

Overview

This chapter presents the data relative to each research question in both written and visual presentation format (Tables 9 through 10 and Figures 2 through 19). First, descriptive statistics are presented regarding the relevant variables. Second, for each research question, multiple statistical analyses are discussed. It is important to note that although main effects are presented, of primary interest are the interaction results. In fact, interpretation and generalization of the main effects are not recommended due to the presence of strong interaction results (Lomax, 2007). The chapter concludes with a summary of the findings from the data analyses.

Descriptive Statistics

See Table 9 for mean percentage implementation level of each of the nine independent variables among the participating schools. Overall, these schools were found to be implementing approximately 55% of the components of ISM as measured by the IET. Their mean implementation on the six Key Features of ISM ranged from 18.88% (SBR) to 72.53% (DBDM). Finally, mean implementation of the behavioral components exceeded implementation of the academic components by over 10%.

Descriptive data also were calculated for the dependent variables. Table 10 provides means and standard deviations on the dependent variable for the first three research questions, disaggregated by implementation level and 2007 student reading level. Descriptive data on the dependent variable ODR will be discussed later due to unique features of these data.
Table 9

*Descriptive Statistics on the Independent Variables*

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IET</td>
<td>2660</td>
<td>55.21</td>
<td>7.73</td>
</tr>
<tr>
<td>AL</td>
<td>2660</td>
<td>41.93</td>
<td>8.92</td>
</tr>
<tr>
<td>CRP</td>
<td>2660</td>
<td>29.67</td>
<td>8.00</td>
</tr>
<tr>
<td>CSP</td>
<td>2660</td>
<td>52.18</td>
<td>9.11</td>
</tr>
<tr>
<td>DBDM</td>
<td>2660</td>
<td>72.53</td>
<td>9.41</td>
</tr>
<tr>
<td>SBR</td>
<td>2660</td>
<td>18.88</td>
<td>11.96</td>
</tr>
<tr>
<td>TT</td>
<td>2660</td>
<td>50.95</td>
<td>10.74</td>
</tr>
<tr>
<td>ACADEM</td>
<td>2660</td>
<td>59.52</td>
<td>22.29</td>
</tr>
<tr>
<td>BEHAV</td>
<td>2660</td>
<td>70.82</td>
<td>10.50</td>
</tr>
</tbody>
</table>

Table 10

*Descriptive Statistics on 2008 DIBELS Z-Scores, Disaggregated by Implementation Level and 2007 Reading Level*

<table>
<thead>
<tr>
<th>Implementation Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007 Reading Level</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>IET</td>
<td>-0.03</td>
<td>0.93</td>
<td>-0.58</td>
<td>0.93</td>
</tr>
<tr>
<td>Medium</td>
<td>0.04</td>
<td>1.07</td>
<td>-0.08</td>
<td>0.82</td>
</tr>
<tr>
<td>High</td>
<td>0.09</td>
<td>1.09</td>
<td>0.31</td>
<td>1.05</td>
</tr>
<tr>
<td>Overall</td>
<td>0.04</td>
<td>1.04</td>
<td>-0.12</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>AL</td>
<td>CRP</td>
<td>CSP</td>
<td>DBDM</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>Low</td>
<td>-0.95</td>
<td>0.93</td>
<td>-0.57</td>
<td>-0.02</td>
</tr>
<tr>
<td>Medium</td>
<td>0.03</td>
<td>1.07</td>
<td>0.55</td>
<td>0.15</td>
</tr>
<tr>
<td>High</td>
<td>0.90</td>
<td>0.74</td>
<td>0.55</td>
<td>0.21</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.11</td>
<td>0.98</td>
<td>0.00</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Research Question One: Are There Significant Differences in Student Academic Outcomes Based on School ISM Implementation Level?

Hypotheses

It was anticipated that significant differences in student academic outcomes would be found based on school ISM implementation level. Specifically, High implementing schools were expected to have higher student reading outcomes than Medium and Low implementing schools. In addition, Medium implementing schools were expected to have higher student reading outcomes than Low implementing schools.
Findings

Results of the factorial ANOVA demonstrated a significant effect of IET implementation level on 2008 student reading outcomes, $F(2, 2651) = 10.308, p=.000, \eta=.008$. Post-hoc analyses suggest that students attending High ($M=0.08$) implementing schools outperformed students attending Medium ($M=-0.12$) and Low ($M=0.03$) implementing schools, although this difference was significant only when considering the difference between High and Medium implementing schools. In addition, students attending Medium implementing schools performed significantly lower than students attending High and Low implementing schools. See Figure 2 for a visual depiction of the average 2008 student reading performance by school IET Implementation Level.

![Figure 2. 2008 Mean Student DIBELS Z-Scores by School IET Implementation Level](image)

Significant differences in student reading outcomes based on overall ISM implementation level were also found to differ based on initial student reading
performance. Specifically, an interaction between school implementation level and student spring 2007 reading performance was found, \( F(4, 2651) = 19.153, p = .000, \eta = .028 \). Students who were in the Low performing group on the spring 2007 assessment and attended High (\( M = 0.09 \)) implementing schools performed above the participant average in 2008, whereas those who attended Low (\( M = -0.03 \)) or Medium (\( M = -0.58 \)) implementing schools performed below the participant average. In contrast, students who were in the Medium performing group on the spring 2007 assessment and attended High (\( M = 0.06 \)) or Low (\( M = 0.04 \)) implementing schools performed above average and those who attended Medium (\( M = -0.08 \)) implementing schools performed below average. Finally, although students who were in the High performing group on the spring 2007 assessment performed above the participant average across all implementation levels in 2008, their relative performance was highest in Medium (\( M = 0.31 \)) implementing schools, middle in the Low (\( M = 0.09 \)) implementing schools, and lowest in High (\( M = 0.08 \)) implementing schools.

For a visual depiction of the mean z-scores disaggregated by implementation level and spring 2007 reading performance, see Figure 3. Because of the unclear and contradictory relationship between implementation level and outcomes for students who initially were Medium performers and those attending Medium implementing schools, an additional figure using only High and Low performers and implementing schools was created to more clearly present these relationships (see Figure 4).
Figure 3. 2008 Student DIBELS Z-Scores by School IET Implementation Level and spring 2007 DIBELS Z-Scores

Figure 4. Figure 3 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment
Research Question Two: Are There Significant Differences in Student Academic Outcomes Based on School Implementation Level of Each of the Six Key Features of Ohio’s ISM?

Hypotheses

It was expected that differences in student reading outcomes would be found based on school implementation level of each of the Six Key Features of Ohio’s ISM. For each feature, it was anticipated that High implementing schools would be found to have higher student reading outcomes than Medium and Low implementing schools. In addition, Medium implementing schools were expected to have higher student reading outcomes than Low implementing schools. However, it was anticipated that differences in the strength of the relationship between implementation level and student reading outcomes would be found. For example, it was anticipated that the proportion of variance in student reading outcomes explained by each of the Six Key Features would differ, and this would speak to the relative importance of each factor related to student outcomes. Given the lack of previous research exploring the relative importance of each Key Feature, there were no hypotheses regarding which would be most significant.

Findings

For this research question, it is important to consider each Key Feature in isolation. See Table 10 for a summary of all data in tabular format. For a visual depiction of the mean student z-score by school implementation level see Figures 5, 8, 11, 14, 17, and 20. For a visual depiction of mean z-score disaggregated by implementation level and spring 2007 reading performance, see Figures 6, 9, 12, 15, 18,
and 21. Because of the unclear and contradictory relationship between implementation level and outcomes for students who initially were Medium performers and those attending Medium implementing schools, additional figures using only High and Low performers and implementing schools were created to more clearly present these relationships (see Figures 7, 10, 13, 16, 19, and 22).

Administrative Leadership

Results of the factorial ANOVA did not support a significant main effect of AL implementation level on student reading outcomes, $F(2, 2651) = 0.487, p=.614, \eta=.000$. However, a significant interaction between school AL implementation level and student spring 2007 reading performance was found, $F(4, 2651) = 53.584, p=.000, \eta=.075$. Students who were in the Low performing group on the spring 2007 assessment and attended High ($M=0.03$) implementing schools performed above the participant average in 2008, whereas those who attended Medium ($M=-0.03$) and Low ($M=-0.95$) implementing schools performed below the participant average. Results for Medium performing students were more contradictory, with those attending Medium ($M=0.04$) implementing schools faring above average and those attending Low ($M=-0.03$) and High ($M=-0.00$) implementing schools faring below average. Finally, students who were in the High performing group on the spring 2007 assessment performed above the participant average across all implementation levels; however, their performance was highest in Low ($M=0.90$) implementing schools followed by the Medium ($M=0.09$) and High ($M=0.02$) implementing schools.
Figure 5. 2008 Mean Student DIBELS Z-Scores by School AL Implementation Level

Figure 6. 2008 Student DIBELS Z-Scores by School AL Implementation Level and spring 2007 DIBELS Z-Scores
The factorial ANOVA found no significant effect of CRP on 2008 student reading outcomes, $F (2, 2651) = 0.403, p=.668, \eta=.000$. However, a significant interaction was found between CRP implementation level and initial student reading performance, $F (4, 2651) = 18.276, p=.000, \eta=.027$. Students whose 2007 DIBELS performance placed them in the Low performing group and attended Medium ($M = 0.04$) implementing schools performed above the participant average in 2008, whereas those who attended Low ($M = -0.03$) and High ($M = -0.45$) implementing schools performed below average. In contrast, students in the Medium performing 2007 group who attended Low ($M = 0.04$) implementing schools performed above the participant average whereas those who attended Medium ($M = -0.01$) and High ($M = -0.00$) implementing schools performed below average. Finally, all students who were initially in the High performing group
Figure 8. 2008 Mean Student DIBELS Z-Scores by School CRP Implementation Level

Figure 9. 2008 Student DIBELS Z-Scores by School CRP Implementation Level and spring 2007 DIBELS Z-Scores
Figure 10. Figure 9 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment.

performed above average on the 2008 DIBELS assessment; however, the highest scores were found in those students attending High (\(M = 0.42\)) implementing schools followed by those attending Low (\(M = 0.09\)) and Medium (\(M = 0.03\)) implementing schools.

Collaborative Strategic Planning

A main effect of CSP implementation level on 2008 student reading outcomes was supported by the factorial ANOVA, \(F(2, 2651) = 8.623, p=0.000, \eta=0.006\). Pairwise comparisons suggested that students attending High (\(M = 0.13\)) implementing schools performed significantly better than students attending both Medium (\(M = -0.07\)) and Low (\(M = -0.00\)) implementing schools.

Student outcomes also were found to differ based on the interaction between CSP implementation level and 2007 student reading performance, \(F(4, 2651) = 32.349,\)
$p=.000$, $\eta=.047$. Students initially in the Low performing group performed most below average when they attended Low ($M = -0.57$) implementing schools, less below average when they attended Medium ($M = -0.06$) implementing schools, and above average when they attended High ($M = 0.17$) implementing schools. Students who were initially Medium and High performers scored below the participant average when they attended Medium ($M = -0.08$ and $M = -0.07$, respectively) implementing schools and above the participating schools when they attended Low ($M = 0.05$ and $M = 0.55$, respectively) and High ($M = 0.06$ and $M = 0.08$, respectively) implementing schools.

*Figure 11.* 2008 Mean Student DIBELS Z-Scores by School CSP Implementation Level
Figure 12. 2008 Student DIBELS Z-Scores by School CSP Implementation Level and spring 2007 DIBELS Z-Scores

Figure 13. Figure 12 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment
Data-based Decision Making

A main effect of DBDM implementation level of student reading outcomes was found, $F (2, 2651) = 12.814, p=.000, \eta=.010$. Pairwise comparisons suggested that High ($M = 0.08$) and Low ($M = 0.12$) implementing schools performed significantly higher than Medium ($M = -0.11$) implementing schools on the outcome variable.

A significant interaction between DBDM implementation level and initial student reading performance also was found, $F (4, 2651) = 11.980, p=.000, \eta=.018$. Students who were initially in the Low performing group and attended Low ($M = -0.02$) implementing or Medium ($M = -0.45$) implementing schools continued to score below the participant average in 2008, whereas their counterparts who attended High ($M = 0.09$) implementing schools scored above the participant average. Students initially in the Medium performing group who attended Low ($M = 0.15$) and High ($M = 0.06$) implementing schools performed above the participant average whereas those attending Medium ($M = -0.08$) implementing schools performed below the participant average. Finally, students initially in the High performing group outperformed the participant average across all levels of implementation; however, their performance was highest in the Medium ($M = 0.22$) implementing schools followed by the Low ($M = 0.21$) implementing and the High ($M = 0.08$) implementing schools, respectively.
Figure 14. 2008 Mean Student DIBELS Z-Scores by School DBDM Implementation Level

Figure 15. 2008 Student DIBELS Z-Scores by School DBDM Implementation Level and spring 2007 DIBELS Z-Scores
Scientifically Based Research

A main effect of SBR implementation level on 2008 student outcomes was evidenced, $F(2, 2651) = 4.453, p=.012, \eta=.003$. Although students attending High ($M = 0.08$) implementing schools outperformed those attending both Low ($M = -0.04$) and Medium ($M = -0.02$), pairwise comparisons suggested these differences only were significant when considering the difference between High and Low implementing schools.

Differences in student reading outcomes when considering the interaction between SBR implementation level and initial student reading performance were found, $F(4, 2651) = 53.289, p=.000, \eta=.074$. Participants with initially Low and Medium reading performance who attended Low ($M = -0.08$ and $M = -0.03$, respectively) and Medium ($M = -0.95$ and $M = -0.03$, respectively) implementing schools scored below the
participant average whereas those attending High ($M = 0.09$ and $M = 0.06$, respectively) implementing schools scored above the participant average. In contrast, initially High performing students who attended Low ($M = -0.00$) implementing schools performed below the participant average whereas those who attended Medium ($M = 0.90$) and High ($M = 0.08$) performing schools scored above the participant average. Interestingly, initially Low performing students attending High ($M = 0.09$) implementing schools scored on average higher than initially High performing and Medium performing students attending Low ($M = -0.03$ and $M = -0.00$, respectively) implementing schools on the 2008 DIBELS measure.

![Figure 17. 2008 Mean Student DIBELS Z-Scores by School SBR Implementation Level](image-url)
Figure 18. 2008 Student DIBELS Z-Scores by School SBR Implementation Level and spring 2007 DIBELS Z-Scores

Figure 19. Figure 18 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment
Results of the factorial ANOVA indicated a significant main effect of TT implementation level on 2008 student reading outcomes, \( F(2, 2651) = 10.308, p = .000, \eta^2 = .008 \). Specifically, pairwise comparisons suggest students attending both High (\( M = 0.08 \)) and Low (\( M = 0.04 \)) implementing schools performed significantly greater than those attending Medium (\( M = -0.12 \)) implementing schools.

An interaction between TT implementation level and initial student performance level was also found, \( F(4, 2651) = 19.153, p = .000, \eta^2 = .028 \). Students who performed in the Low group on the spring 2007 benchmark assessment and attended High (\( M = 0.09 \)) implementing schools performed above the participant average in 2008, whereas their counterparts who attended Medium (\( M = -0.58 \)) and Low (\( M = -0.03 \)) implementing schools remained below the participant average in 2008. For students initially performing within the Medium range, those who attended Medium (\( M = -0.08 \)) implementing schools performed below the participant average in 2008 whereas those who attended Low (\( M = 0.04 \)) and High (\( M = 0.06 \)) implementing schools performed above the participant average in 2008. Finally, initially High performing students performed above the participant average across all implementation levels, although their relative performance was highest in Medium (\( M = 0.31 \)) implementing schools followed by Low (\( M = 0.09 \)) and High (\( M = 0.08 \)) implementing schools.
Figure 20. 2008 Mean Student DIBELS Z-Scores by School TT Implementation Level

Figure 21. 2008 Student DIBELS Z-Scores by School TT Implementation Level and spring 2007 DIBELS Z-Scores
Figure 22. Figure 21 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment

Comparison

The relative importance of implementation level of each of the six Key Features varied as evidenced by the proportion of variance in student reading outcomes explained by each. Specifically, when considering the interaction between implementation level and initial student reading performance, the Key Features having the largest effect were (in descending order of importance): AL ($\eta = .075$), SBR ($\eta = .074$), CSP ($\eta = .047$), TT ($\eta = .028$), CRP ($\eta = .027$), and DBDM ($\eta = .018$).

Research Question Three: Are There Significant Differences in Student Academic Outcomes Based on School Implementation Level of the (a) Academic Components of ISM and (b) Behavioral Components of ISM?
**Hypotheses**

It was expected that differences in student academic outcomes would be found based on school implementation level of the academic and behavioral components. For both components, it was anticipated that High implementing schools would be found to have higher student academic outcomes than Medium and Low implementing schools. In addition, Medium implementing schools would be found to have higher student academic outcomes than Low implementing schools. However, it was anticipated that differences in the strength of the relationship between implementation level and student academic outcomes would be found. Specifically, the proportion of variance in student academic outcomes explained by the academic and behavior components was anticipated to differ, with a greater proportion of variance explained by the academic components. In addition, it was expected that the proportion of variance explained by the integrated model (research question 1) would be greater than that explained by either the academic or behavioral components in isolation.

**Findings**

**Academic**

Results of the factorial ANOVA indicate implementation level of the academic components of the IET significantly predict 2008 student reading outcomes, $F(2, 2651) = 8.367, p=.000, \eta=.006$. Pairwise comparisons suggest students attending High ($M = 0.08$) implementing schools performed significantly better than students attending Medium ($M =-0.02$) implementing and Low ($M =-0.12$) implementing schools.
Additionally, students attending Medium ($M = 0.08$) implementing schools performed significantly better than students attending Low ($M = 0.08$) implementing schools.

An interaction between implementation level and initial student reading performance also was found, $F (4, 2651) = 18.920, p=.000, \eta=.028$. Specifically, students who performed in the Low group on the 2007 DIBELS assessment and attended Low ($M = -0.15$) or Medium ($M = -0.44$) implementing schools performed below the participant average in 2008; whereas initially Low performers who attended High ($M = 0.09$) implementing schools performed above the participant average. Students initially performing in the Medium and High groups evidenced a similar pattern. Namely, those who attended Low ($M = -0.14$ and $M = -0.13$, respectively) implementing schools performed below the participant average in 2008 whereas those who attended Medium ($M = 0.01$ and $M = 0.38$, respectively) or High ($M = 0.06$ and $M = 0.08$, respectively) implementing schools performed above the participant average. Interestingly, initially Low performing students who attended High ($M = 0.09$) implementing schools performed higher in 2008 than initially High performing students who attended High ($M = 0.08$) implementing schools.

**Behavioral**

A significant effect of implementation level of the behavior components of the IET on 2008 student reading outcomes was found, $F (2, 2651) = 8.965, p=.000, \eta=.007$. However, the direction of this relationship was not in the expected direction. Specifically, pairwise comparisons indicated that students attending High ($M = -0.08$) implementing schools evidenced significantly lower 2008 reading performance than
Figure 23. 2008 Mean Student DIBELS Z-Scores by School ACADEM Implementation Level

Figure 24. 2008 Student DIBELS Z-Scores by School ACADEM Implementation Level and spring 2007 DIBELS Z-Scores
Figure 25. Figure 24 modified to exclude Medium implementers and Medium performers on the spring 2007 DIBELS assessment

students attending Medium ($M = 0.13$) and Low ($M = 0.04$) implementing schools. Similar to the academic components, an interaction between behavior implementation level and initial student reading performance was found, $F(4, 2651) = 8.687, \ p=.000, \ \eta=.013$. Interestingly, among initially Low performing students, those who attended High ($M = -0.36$) implementing schools performed the lowest of any group and those who attended Low ($M = -0.03$) implementing schools also performed below the participant average; in contrast, those attending Medium ($M = 0.17$) implementing schools performed above the participant average. Among initially Medium performing students, those attending Low ($M = 0.04$) and Medium ($M = 0.07$) implementing schools performed above the participant average, whereas those attending High ($M = -0.05$) implementing schools performed below the participant average. Finally, among initially
Figure 26. 2008 Mean Student DIBELS Z-Scores by School BEHAV Implementation Level

Figure 27. 2008 Student DIBELS Z-Scores by School BEHAV Implementation Level and spring 2007 DIBELS Z-Scores
High performing students all performed above the participant average; however, this effect decreased when considering those who attended Low ($M = 0.09$), Medium ($M = 0.13$), and High ($M = 0.21$) implementing schools, respectively.

**Comparison**

Implementation of the academic components of ISM explained more variability in reading outcomes than did implementation of the behavioral components ($\eta = .028$ and $\eta = .013$, respectively). Overall implementation of ISM ($\eta = .028$, as presented in the first research question) did not explain any additional variation above and beyond implementation of the academic components.
Research Question Four: Are There Significant Differences in School-level Behavioral Outcomes Based on School-level ISM Implementation Level?

Hypotheses

Significant differences in school-level behavioral outcomes based on school-level ISM implementation level were expected. Specifically, it was anticipated that High implementing schools would have more favorable school-level behavioral outcomes than Medium implementing schools and Medium implementing schools would have more favorable school-level behavioral outcomes than Low implementing schools.

Findings

As previously discussed in the Method section, ODR data was available for only two of the six participating schools. As a result, the anticipated statistical analyses could not be computed and behavioral data were examined only descriptively. To protect the identity of the participating schools, ODRs per 100 students will not be revealed. Instead, the ratio of ODRs from 2006-2007 to 2007-2008 will be used as a proxy for improvement.

The Higher implementing school evidenced a larger number of ODRs than the Lower implementing school. In addition, although both schools decreased in the number of ODRs per 100 students from 2006-2007 to 2007-2008, the Lower implementing school evidenced a larger percentage decrease in ODRs (i.e., 43%) during this time period than the Higher implementing school (i.e., 31%).
CHAPTER V: DISCUSSION

Introduction

This chapter provides a discussion of the findings, limitations, delimitations, and implications for research and practice. It is anticipated that together these components will provide an integrated view of the study’s relevance.

Discussion of the Findings

To best discuss the study’s findings in light of their meaning, relevance, and relationship to other research, it is important to draw attention to factors that shaped the presentation of the results. First, the results of Medium implementing schools and initially Medium performing students often resulted in a clouding of the substantial interaction results found between the Low and High groups. In many cases, clear and consistent patterns emerged for initially High and Low performing students as well as High and Low implementing schools; in contrast, results were more variable and less significant when considering initially Medium performing students and Medium implementing schools.

Additional examination of the descriptive statistics and overall findings suggested that (a) Medium implementing were more similar to Low implementing schools than High implementing schools with respect to implementation levels, which may partially explain suboptimal outcomes for Medium implementing schools across many of the independent variables, and (b) initially Medium performing students may not be as impacted by implementation levels as other groups. Supporting the latter assertion, the mean performance of initially Medium performing students was often strikingly close to
the participant mean \((z = 0)\) across implementation levels. As a result of the weak and inconsistent effect in these middle groups – and in order to enhance clarity and focus on the most critical results – this discussion will focus exclusively on High and Low implementing schools and High and Low initially performing students when interpreting the interaction results (main effect results, due to their relative simplicity, will include Medium populations).

A second issue to consider is that the dependent variable was assessed using z-scores. As a result, a negative outcome value indicates only that the group mean was below the participant average. Therefore, a negative 2008 z-score among 2007 High performing students does not mean that the group actually decreased in DIBELS score from 2007 to 2008; rather, it indicates that their mean rank position when compared to all participants decreased (because their initial High status indicates that their initial mean rank was a positive value). In fact, to prevent such misinterpretation, analyses of descriptive data for the actual ORF scores (rather than the z-scores) were completed.

Results suggest that across all implementation types and initial student performance levels mean scores on the DIBELS ORF assessments did increase from 2007 to 2008. What may have decreased for some populations is their relative rank standing when compared to the rest of the sample due to more substantial increases in other groups. Therefore, it is clear that regardless of implementation level or initial student performance, all levels of the independent variables did evidence mean improvements in oral reading fluency performance from 2007 to 2008.
Finally, it is important to note that the interaction results are of primary interest in this study. Main effect results are presented simply because (a) main effects are typically presented in conjunction with interaction effects, and (b) it was decided a priori to report both sets of effects. In the current study, however, the primary interest is whether or not differences in outcomes based on implementation level were found when considering initial student performance. Because it was revealed that differences do exist based on initial student performance, the main effects – which assume the same effect of implementation level on student outcomes across all groups – lose much of their meaning (Lomax, 2007). As a result, extreme caution is recommended when interpreting the main effects and a focus on the interaction effects is advised.

*Academic Outcomes Based on Overall Implementation Level*

Consistent with the original hypotheses, results of the factorial ANOVA found a significant effect of ISM implementation level on student reading outcomes. Specifically, students attending High implementing schools performed higher than students attending Medium and Low implementing schools, although this difference was statistically significant only when considering Medium implementing schools. In addition, students attending Medium implementing schools performed significantly lower than students attending High and Low implementing schools.

A statistically significant interaction between ISM school implementation level and student spring 2007 reading performance also was found. Specifically, initially Low performing students performed above average when they attended High implementing schools and below average when they attended Low implementing schools. In contrast,
initially High performing students outperformed the average across Low and High implementation levels, although their performance was better in Low implementing schools than in High implementing schools.

Overall, these results suggest that school implementation level does have a significant impact on reading performance. This is consistent with prior research supporting an effect of comprehensive school reform implementation on student outcomes (e.g., Datnow et al., 2000; Nunnery, 1997; the Urban Institute, 2007). Although the current finding that students attending High implementing schools outperformed those in Medium and Low implementing schools confirmed initial hypotheses, the finding that students attending Medium implementing schools significantly underperformed students attending Low implementing schools did not. Specifically, it was expected that students attending Medium implementing schools would perform significantly better than those attending Low implementing schools. The reason for this discrepancy in findings can only be speculated; however, as previously stated, the findings for the Medium groups in general were weak and inconsistent and main effects should be interpreted cautiously.

More importantly, findings suggest that the impact of implementation level is moderated by initial student performance. Interestingly, initially Low performing students who attended High implementing schools performed equally to initially High performing students attending Low implementing schools, and outperformed initially High performing students attending High implementing schools. That is, the mean achievement gap between Low and High performing students was eliminated in High
implementing schools. These represent the first known results to examine the relationship of ISM implementation on different achievement groups. The results are quite promising for schools seeking to have all students meet basic academic competencies, consistent with the mandates of legislation such as NCLB (2001). This is particularly true given the vast improvements evidenced by Low performing students, coupled with continued improvements in raw outcomes among other students (despite their change in relative rank standing).

Although statistical significance was found across analyses for this research question, it is important also to consider measures of practical significance such as effect sizes (e.g., Kirk, 1996; Sink & Stroh, 2006). It is widely accepted that for partial eta squared, a small effect is 0.01 to 0.06, medium effect is 0.06 to 0.14, and a large effect is 0.14 and higher (Sink & Stroh). Based on these guidelines, the observed effect sizes for the main and interaction effects are considered small.

**Academic Outcomes Based on Implementation Level of the Six Key Features**

A main effect of implementation level on student reading outcomes was found when considering four of the Six Key Features of Ohio’s ISM: Collaborative Strategic Planning (CSP), Data-Based Decision Making (DBDM), Scientifically-Based Research (SBR), and Academic and Behavioral Supports Across Three Tiers (TT). For CSP, students attending High implementing schools performed significantly better than students attending Low and Medium implementing schools. For DBDM and TT, students attending High and Low implementing schools performed significantly better than students attending Medium implementing schools. For SBR, students attending
High implementing schools performed significantly better than students attending Low implementing schools. Across all four of these, students attending High implementing schools significantly outperformed students attending at least one other school type. Main effects were not found on the remaining two Key Features: Administrative Leadership (AL) and Culturally Responsive Practices (CRP). For AL, it is likely that a significant main effect was not found because the strength of the interaction effect diminished the main effect. It is unknown why a main effect was not found for CRP; however, it is notable that although not significant, results were not in the expected direction on this Key Feature (i.e., students attending Low implementing schools performed above average and those attending High implementing schools performed below average).

In addition, a significant interaction between implementation level and 2007 student reading performance was found across each of the Six Key Features of Ohio’s ISM. For AL, CSP, DBDM, and TT, initially Low performing students who attended Low implementing schools performed below average whereas those attending High implementing schools performed above average; in contrast, initially High performing students performed above average across implementation levels, although the highest performance was found in Low implementing schools. For SBR, both initial performance groups (i.e., Low and High) evidenced below average performance in Low implementing schools and above average performance in High implementing schools. However, initially Low performing students underperformed initially High performing students in the Low implementing schools and outperformed initially High performing
students in the High implementing schools. For CRP, initially Low performing students performed below average at both Low and High implementing schools, although their mean performance was lower at High performing schools; similarly, initially High performing students performed above average at both Low and High implementing schools, although their mean performance was higher at High implementing schools.

Two notable findings emerged from these analyses. As expected, initially Low performing students who attended Low implementing schools performed below initially High performing students who attended Low implementing schools across all nine independent variables. However, for seven of the nine independent variables (i.e., IET, AL, CSP, DBDM, SBR, TT, and ACADEM), initially Low performing students who attended High implementing schools outperformed their initially High performing counterparts at the same schools. Second, on the variables AL, BEHAV, CRP, CSP, DBDM, IET, TT a disordinal interaction – a more dramatic type of interaction – was found (Stevens, 1999). This type of interaction indicates that Low implementation was superior for one group of initial implementers whereas High implementation was superior for another group of initial implementers (Stevens, 1999). This interaction was in the expected direction for the variables AL, CSP, DBDM, IET, and TT.

It is important also to consider the proportion of variability explained in student reading outcomes by each of the six Key Features using the previously described effect size guidelines (i.e., 0.01 to 0.06 = small effect; 0.06 to 0.14 = medium effect; 0.14 and higher = large effect; Sink & Stroh, 2006). Based on these guidelines, many of the observed effects for this research question were small. However, the interaction effects
between initial student performance and implementation level represented a medium
effect for the variables AL and SBR. This suggests that these variables have a more
substantial effect on student outcomes than the others.

*Academic Outcomes Based on Academic and Behavioral Implementation Level*

A main effect of implementation level for both the academic and behavioral
components of ISM was found. However, the direction of the effect differed by
component. For the academic components, students attending High implementing
schools significantly outperformed those implementing Medium and Low implementing
schools, and students attending Medium implementing schools significantly
outperformed those attending Low implementing schools. This pattern of findings
supported the initial hypotheses. In contrast, the findings for the behavioral component
were not in the expected direction. Specifically, students attending High implementing
schools significantly underperformed those attending Medium and Low implementing
schools.

An interaction between implementation level and initial student reading
performance was also found for both the academic and behavioral components. For the
academic components, results supported the initial hypotheses. Specifically, initially
Low performing students who attended Low implementing schools performed below
average, whereas their counterparts attending High implementing schools performed
above average. Initially High performing students who attended Low implementing
schools also performed below the participant average although those attending High
implementing schools performed above average. Interestingly, initially Low performing
students who attended High implementing schools performed higher in 2008 than initially High performing students who attended the same schools. For the behavioral components, results were not in the expected direction. Specifically, initially Low performing students who attended High and Low implementing schools performed below the participant average, although those attending High implementing schools performed the lowest; in contrast, initially High performing students performed above average across both implementation levels, although this effect decreased when considering those who attended Low and High implementing schools, respectively.

Although effect sizes were considered small across both independent variables, implementation of the academic components of ISM explained more variability in reading outcomes than did implementation of the behavioral components. This is not unexpected given previous research suggesting the implementation of academic models of support has a stronger effect on academic outcomes than the implementation of behavioral models of support (Stewart et al., 2007). Interestingly, overall implementation of ISM did not explain any additional variation of reading outcomes beyond implementation of the academic components. This is surprising given previous research suggesting integrated models of support produce more significant academic outcomes than academic models of support (Stewart et al.). However, an important distinction between the Stewart et al. study and the current study may partially explain this difference. In the Stewart et al. study, schools were either implementing (a) academic only, (b) behavior only, or (c) integrated models of support. In this study, all schools
were implementing integrated models of support, albeit to different degrees of implementation for the different components.

**Behavioral Outcomes Based on Overall Implementation Level**

Statistical analyses could not be computed due to the small number of schools with available behavioral data. However, descriptive examination of the behavioral data suggested the Higher implementing school evidenced (a) a larger number of ODRs, and (b) a smaller percentage decrease in ODRs from 2006-2007 to 2007-2008 than the Lower implementing school.

Unfortunately, the extremely small sample size limits the conclusions that can be drawn from these data. Such analyses raise more questions than they answer and simply indicate the need for further research using larger samples. The results should particularly be viewed with caution given previous research documenting a positive effect of ISM on behavioral outcomes (Stewart et al., 2007). Overall, no implications can be drawn from these analyses.

**Limitations**

There are several limitations associated with the current study. First, the use of a causal-comparative design limits the degree to which one can assume a causal relationship between ISM implementation and student outcomes. It seems plausible that increased ISM implementation would cause increased student performance; however, an alternative explanation for the relationship between ISM implementation and student outcomes is that schools that have higher performing students to begin with have the time and resources to implement ISM at a higher level. In order to make causal attributions,
schools would need to be randomly selected and randomly assigned to the implementation levels or changes in implementation and student performance would need to be examined longitudinally, designs that were not feasible for the study. The effects of this limitation, however, have been reduced using two procedures. First, the sample was made more homogeneous using t-tests, thereby reducing pre-existing differences. Second, a factorial design was used to examine the interaction between implementation level and 2007 student reading performance. Therefore, differences in initial student performance were incorporated into the analyses.

Second, the sample of schools used in the study is potentially concerning. Because the independent variables are measured on the school-level (rather than student level), the small number of schools participating in the study results in limited diversity in the independent variable. A larger number of schools would add to the robustness of the analyses. However, this was not feasible given: (a) the limited number of schools who were administered the IET, and (b) a lack of resources for administering the time and labor-intensive IET to additional schools. In addition, the degree to which the sample reflected all schools implementing ISM is unknown. For example, it is unknown if the invited schools that chose not to participate differed in meaningful ways from the invited schools that chose to participate. In addition, schools that had the IET administered may have been more invested in the ISM process than ISM schools that did not have the IET administered. This homogeneous sample also might evidence different findings than a heterogeneous sample including more diverse schools (e.g., urban schools). Finally, all participating schools received some support for ISM (e.g., financial,
consultative). Therefore, it is unknown if the results would generalize to schools who are attempting to implement ISM without these supports. Together, these concerns limit the external validity of the study.

Third, concerns regarding the internal validity of the study exist. For example, the validity of the IET has remained largely unexplored. Therefore, the degree to which it accurately differentiates between High, Medium, and Low implementing schools is unknown. In addition, because the more advanced items (i.e., those unique to Phase III) were not directly assessed on earlier Phases, it is possible that the performance of Phase II schools was artificially inflated because percentage scores were used (i.e., it is possible a Phase II schools may have scored a higher percentage than a Phase III school despite not having implemented as many components). Despite this potential limitation, the percentage scores remained the most appropriate measurement option. Finally, although the items on the IET were deemed to theoretically load onto the six Key Features of Ohio’s ISM, this was not confirmed through a factor analysis. As a result, the degree to which scores on each of the Key Feature scales accurately reflect implementation of that Key Feature can only be assumed.

A final potential limitation of the study concerns whether the interaction findings might reflect regression towards the mean rather than true differences between groups (the main effect findings would not be impacted by regression to the mean). Regression towards the mean is a term used to describe the phenomenon that a variable initially measured to be extreme will tend to fall closer to the center of the distribution on later measurements (Davis, 1976). This is because the more extreme a participant scores on an
initial measurement, the less extreme the error component is likely to be on further measures, thus bringing the observed score closer to the mean (Shaughnessy & Zechmeister, 1990). Although regression toward the mean cannot be prevented, it can be minimized and compensated for.

Streiner (2001) suggests two techniques for doing so. First, an extreme group should never be selected on the basis of a single test score. When a single score is used for group selection, there is an increased likelihood that a participant is assigned to the extreme group because his or her observed score is above the criterion while the true score is below it. In the current study, the outcome variable represented the z-score associated with the median DIBELS ORF performance on three ORF assessments; therefore, reliance on multiple measures was used to reduce the potential for regression toward the mean. In addition, regression toward the mean can be compensated for using a control group that had similar pre-test scores but was not given the intervention. Because both groups will be affected equally by regression to the mean, the resulting between-group differences will indicate the degree of effect (Streiner). Although a control group was not available in the study, it is nonetheless plausible to assume that regression to the mean would affect all three groups (Low, Medium, and High implementing schools) equally. That is, initially Low achieving students would demonstrate this effect whether they attended Low, Medium, or High implementing schools. What was found, however, was that the effect was not universal across the school types. Rather, finding suggested that these groups were affected quite differentially, with Low achieving students attending Low implementing schools
typically remaining below average and Low achieving students attending High implementing schools often surpassing the overall participant mean.

Delimitations

This study also has several delimitations. First, the study was limited to districts that completed the IET. Given the time and resources required to administer the IET, it simply would not be feasible to either: (a) administer this instrument to a random sample of schools implementing ISM, or (b) create and validate a new instrument for assessing ISM implementation.

Also, statistical methods were used to obtain a more homogeneous final sample from the entire available sample of schools. This resulted in a sample of non-urban, demographically similar schools that are less diverse and higher performing that state averages. A more homogeneous sample was sought because not enough schools were available to generalize results across widely diverse school types; by making the sample more homogeneous, the potential to generalize to similar school types was enhanced.

Finally, the student academic outcomes were limited to DIBELS ORF measures because these measures are sensitive to change, used in all the participating districts, and are frequently used for assessing the effectiveness of instruction and intervention within schools. In addition, DIBELS ORF scores have been found to be significantly correlated with performance on high stakes achievement tests in Ohio, suggesting their practical significance (Vander Meer et al., 2005). The behavioral outcomes were limited to ODRs because they have been determined to be an appropriate metric for monitoring the effectiveness of schoolwide behavioral change programs (Irvin et al., 2004). In addition,
alternative measures (e.g., direct observation, rating scale data) were not feasible for the proposed study.

Implications

Implications for Practice

Given the aforementioned limitations, implications for practice must be stated and interpreted with caution. However, in the absence of other research examining these issues, it seems wise to apply relevant results until additional research confirms or disconfirms their utility. Specifically, a few key implications for practice appear reasonable at this time.

First, an increased emphasis on enhancing ISM implementation is warranted. This study demonstrated that, with a mean overall IET score of 55%, overall implementation of ISM may be less than desirable. An optimal score on the IET is not specified in the manual (Implementation Evaluation Tool Manual, 2007); however, the goal on the SchoolWide Evaluation Tool (SET; Sugai et al., 2001) – an instrument on which the IET is based – is 80% implementation. The observed suboptimal implementation level of 55% is not completely surprising given other research suggesting low levels of treatment integrity for student-level interventions (e.g., Wickstrom et al., 1998) and school-level comprehensive change initiatives (e.g., Datnow et al., 2000). It is important to note that positive effects of ISM implementation level on academic outcomes were found even when considering these suboptimal implementation levels. However, given current findings suggesting implementation levels may impact student outcomes – coupled with similar findings from research on other CSR initiatives (e.g.,
Datnow et al., 2000; Nunnery et al., 1997; Urban Institute, 2007) – a renewed interest in enhancing implementation integrity in order to more strongly improve student outcomes appears warranted. Specifically, before deciding to implement ISM, schools need to create a plan for ensuring implementation is carefully monitored and addressed throughout the longevity of an ISM initiative. In addition, it is important for schools to examine the relationship between implementation and student outcomes. Because high levels of implementation were uniformly associated with higher outcomes, it seems logical for schools to aim for meeting or exceeding a minimum implementation level of 80%, a value (a) higher than the high implementing schools in this study, and (b) recommended by the SET (Sugai et al., 2001).

Although overall IET implementation levels may have been less than ideal, it is important to note that mean IET performance on each of the six Key Features and the reading and behavior components varied widely. This is interesting considering the relatively homogeneous sample. Implementation levels ranged from approximately 19% for SBR to approximately 73% for DBDM. The reasons for such variability are unclear. It could be that some of the features are inherently easier or more difficult to implement within schools’ existing cultural, legal, and financial frameworks. Alternatively, it could be that the training and guidance received by these schools from internal or external consultants focused more specifically on the development of some components over others. Despite the reasons, as a result of this variability, it is important for schools not only to monitor overall implementation levels, but also levels of implementation on the key components of ISM. Although different states and school districts may have a
different number or type of core components of ISM, the six Key Features serve as a
good starting point for those seeking to define and monitor the components of the model.

In addition to exhibiting widely varying implementation levels on the six Key
Features of ISM, it appears that each feature has a differential impact on student reading
outcomes. Specifically, AL and SBR clearly surpass the other features in the amount of
variation in student outcomes explained when considering the interaction results. This
suggests the critical nature of (a) providing a consistent message regarding assumptions
about educating students, and (b) reliance on educational practices and interventions that
have empirical evidence to support their effectiveness (Stollar et al., 2006). The
importance of these two features makes intuitive sense, as they would likely establish a
solid foundation upon which the other features could be developed. For example,
without solid administrative leadership, it is unlikely that (a) broad buy-in from teachers
for the other components, or (b) support for professional development to address the
other key features will be established. As discussed by Gansle and Noell (2007), schools
beginning implementation of a program similar to ISM often find it infeasible to focus on
all aspects of implementation simultaneously. As a result, schools may want to focus
their earliest planning efforts primarily on developing solid and supportive administrative
backing for the initiative as well as researching evidence-based educational practices.
Once these components are established, implementation planning can focus on other
areas of the initiative.

Finally, results of this study suggest that if a school is seeking to improve student
academic outcomes only, an initial focus on implementing the academic components may
be justifiable if resources are not available to simultaneously implement academic and behavioral components. However, given the limitations of the study, it would be unwise to suggest abandoning the behavioral components of ISM. Prior research has clearly demonstrated an effect of behavioral models of support on student academic and behavioral outcomes (Stewart et al., 2007). It is feasible that the unique limitations of the current study resulted in a failure to detect a similar effect. For example, it is possible that initially low levels of problem behaviors among the homogeneous sample resulted in an artificially reduced effect of implementation of the behavioral components of ISM on student outcomes.

**Implications for Research**

Given the previously discussed limitations and delimitations, several avenues for future research also appear warranted.

First and foremost, results of the study suggest that previous methods of evaluating the effectiveness of RTI, SWPBS, or ISM initiatives may be flawed in several ways. First, most previous research examines the effects of implementation versus non-implementation on student outcomes. Many of these studies demonstrate that schools implementing a change initiative perform better on outcomes than non-implementing schools. Most likely, if the current study was designed in this manner, significant results also would have been found when comparing all participating ISM schools with randomly selected comparison schools. However, because ISM implementation levels vary widely, it is important to examine *degree* of implementation on student outcomes. As discovered in this study, Low, Medium, and High implementers evidenced different
outcomes across the research questions. By aggregating these three groups into one group of implementers, these differential effects and nuances would be lost. Future research may want to go beyond dividing implementers into Low, Medium, and High groups and actually use an interval level variable (i.e., percentage implementation) as a predictor variable for student outcomes. Alternatively, future research may want to include a comparison population of non-implementers in addition to examining the different levels of implementation.

Perhaps a more novel implication of the study on research is that in contrast to traditional methods of looking at outcomes for the entire student population, results of this study suggest that researchers may want to examine the differential effects of ISM on different student populations. When outcomes are aggregated for all students, effects may appear diluted because weak effects within some groups may counteract strong effects within other groups. It is recommended that future research examine the effects of ISM on different student populations (e.g., based on initial achievement differences, demographic characteristics, grade-level, etc.). However, special care must be taken before conducting these analyses to ensure that regression to the mean does not explain the findings.

Future research should also more closely examine Medium implementing schools and Medium performing students. It appears from this study that the most consistent results were obtained from High and Low implementing schools and performing students. More variability was found among the middle-of-the-road groups. The reasons for this and the unique characteristics require further examination.
Additional research is also needed to explore the effect of *student-level ISM implementation* on student outcomes. Although ISM this study was primarily concerned with the systemic implementation integrity of a comprehensive ISM program, it is also necessary to evaluate the outcomes of treatment integrity level for individual students in Tiers II and III. This type of research would be similar to that of Telzrow et al. (2000) in that implementation is assessed at the student-level; however, it would be unique because of its focus on ISM rather than IBS.

Finally, it is clear that the impact of ISM implementation variables on behavioral outcomes could not sufficiently be determined due to limitations of the study. It is recommended that future research more closely examine this issue using a larger sample and student-level ODR data.
## Key Features

**Date**________________________

**School**____________________________________________

**City**______________________________

**District__________________________________________**

**Data Collector____________________________**

<table>
<thead>
<tr>
<th>Features</th>
<th>Evaluation Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AL</strong></td>
<td>1. There is a representative building leadership team. The building leadership team lacks diverse representation.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Data Source</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>There is a representative building leadership team. The building leadership team lacks diverse representation.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Meeting notes with a list of members present</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>The building leadership team meets regularly to guide the action planning process. No evidence that the building leadership team meets regularly.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Meeting schedule</strong></td>
</tr>
<tr>
<td><strong>AL</strong></td>
<td>The principal is a member of the building leadership team and regularly attends meetings. The principal does not attend.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Meeting notes with a list of members present</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>The principal occasionally attends meetings. The principal regularly attends meetings (at least 90%).</td>
<td></td>
<td></td>
<td></td>
<td><strong>Principal interview</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>The building leadership team uses the Collaborative Strategic Planning Process to develop an action plan that includes tiered systems of support. No evidence of the Collaborative Strategic Planning process documented on action plan.</td>
<td></td>
<td></td>
<td></td>
<td><strong>CSP Worksheet/Action Plan (must be based on school-wide data)</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>Some evidence of the Collaborative Strategic Planning process documented on action plan. Clear documentation of each step of the Collaborative Strategic Planning process is documented on action plan.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Analysis Guides</strong></td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>Clear documentation of each step of the Collaborative Strategic Planning process is documented on action plan.</td>
<td></td>
<td></td>
<td></td>
<td><strong>Meeting notes</strong></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
</tr>
<tr>
<td>CRP DBDM</td>
<td>5. As part of collaborative strategic planning, data on suspension/expulsion and achievement (test scores, DIBELS, CBM, SWIS) are disaggregated by NCLB subgroups and used in decision-making.</td>
<td>Data are not disaggregated.</td>
<td>Either suspension/expulsion or achievement data analyzed but not both, OR data are disaggregated but there is no evidence that they were used in decision making (i.e. not reflected in the Action Plan).</td>
<td>Data are disaggregated, and the data are used for decision making in the Collaborative Strategic Planning process and are reflected on the Action Plan.</td>
<td>Disaggregated discipline &amp; achievement data (achievement tests, DIBELS, SWIS) Action Plan/CSP Worksheet</td>
</tr>
<tr>
<td>SBR CRP AL</td>
<td>6. Written components of instruction and intervention (instructional integrity checklist or script) are used to ensure criteria for accurate implementation are met and used by those who deliver the instruction.</td>
<td>Instruction and intervention not written down or criteria are not met.</td>
<td>Some but not all instruction is captured in a written script or instructional integrity checklist.</td>
<td>Detailed instruction or intervention plans are written down (scripted) and criteria for accurate implementation are indicated and used by instructors.</td>
<td>Written description of instruction &amp; intervention Instructional integrity checklist</td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>DBDM</strong></td>
<td>7. For all schools with classrooms from Kg-6th grade there is a school-wide electronic (universal) data system in place for academic screening across all grades (DIBELS, AIMSWEB) and all students are screened at least 3 times per year.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No system in place in up to 49% of the classrooms.</td>
<td>System is in place and being used by at least 50% of classrooms.</td>
<td>100% of classrooms screen with data at least 3 times per year.</td>
<td></td>
<td>DIBELS &amp; AIMSWEB reports</td>
</tr>
<tr>
<td><strong>DBDM</strong></td>
<td>8. A school-wide electronic behavioral reporting system (SWIS or an equivalent) to monitor discipline referral data is in place and consistently implemented.</td>
<td>No system in place.</td>
<td>System has been adopted and implementation has begun.</td>
<td>System is in place and entries are no more than 2 weeks behind.</td>
<td>SWIS reports or Equivalent Copy of discipline referral form</td>
</tr>
<tr>
<td><strong>3T</strong></td>
<td>9. There is a written continuum of responses for dealing with specific behavior violation.</td>
<td>No</td>
<td>A listing of possible responses exists but there is no clear written continuum.</td>
<td>Yes, a written continuum with clear steps exists.</td>
<td>Staff Handbook Discipline code – Code of conduct</td>
</tr>
</tbody>
</table>
APPENDIX B:

IMPLEMENTATION EVALUATION TOOL PHASE II
### Key Features

<table>
<thead>
<tr>
<th>Features</th>
<th>Evaluation Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Data Source</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AL</strong></td>
<td>1. There is a representative building leadership team.</td>
<td>The building leadership team lacks diverse representation.</td>
<td>The building leadership team has diverse representation, however, does not include parent membership.</td>
<td>The building leadership team has diverse representation (administration, general education, intervention specialists, etc.) including parent membership.</td>
<td>Meeting notes with a list of members present</td>
<td>0-2</td>
</tr>
<tr>
<td><strong>CSP</strong></td>
<td>2. The building leadership team meets regularly to guide the action planning process.</td>
<td>No evidence that the building leadership team meets regularly.</td>
<td>The building leadership team meets 3 times or fewer per year.</td>
<td>The building leadership team meets regularly (at least 4 times per year).</td>
<td>Meeting schedule</td>
<td></td>
</tr>
<tr>
<td><strong>AL</strong></td>
<td>3. The leadership team has a plan for supporting (on boarding) new team members in learning the components of the model and implementation activities.</td>
<td>No evidence of a plan for supporting new team members in learning the components and activities of the model.</td>
<td>Some evidence of a plan for supporting new team members in learning the components and activities of the model reported or documented.</td>
<td>Clear documentation of action steps for supporting new team members in learning the components and activities of the model.</td>
<td>Written documentation of the plan</td>
<td></td>
</tr>
<tr>
<td><strong>AL</strong></td>
<td>4. The principal is a member of the building leadership team and regularly attends meetings.</td>
<td>The principal does not attend.</td>
<td>The principal occasionally attends meetings.</td>
<td>The principal regularly attends meetings (at least 90%).</td>
<td>Meeting notes with a list of members present Principal interview</td>
<td></td>
</tr>
<tr>
<td><strong>AL</strong></td>
<td>5. The principal demonstrates understanding of the Key Features of ISM.</td>
<td>The principal does not understand the Key Features of ISM.</td>
<td>The principal can talk knowledgeably about some Key Features of ISM, but not all of them.</td>
<td>The principal speaks knowledgeably about the Key Features of ISM.</td>
<td>Principal Interview</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>CSP</td>
<td>6. The building leadership team uses the Collaborative Strategic Planning Process to develop a building action plan that includes tiered systems of support.</td>
<td>No evidence of the Collaborative Strategic Planning process documented on action plan.</td>
<td>Some evidence of the Collaborative Strategic Planning process documented on action plan.</td>
<td>Clear documentation of each step of the Collaborative Strategic Planning process is documented on action plan.</td>
<td>CSP Worksheet/Action Plan (must be based on school-wide data) Analysis Guides Meeting notes</td>
<td>0-2</td>
</tr>
<tr>
<td>CRP DBDM</td>
<td>7. As part of collaborative strategic planning, data on suspension/expulsion and achievement (test scores, DIBELS, CBM, SWIS) are disaggregated by NCLB subgroups and used in decision-making.</td>
<td>Data are not disaggregated.</td>
<td>Either suspension/expulsion or achievement data analyzed but not both, OR data are disaggregated but there is no evidence that they were used in decision making (i.e. not reflected in the Action Plan).</td>
<td>Data are disaggregated, and the data are used for decision making in the Collaborative Strategic Planning process and are reflected on the Action Plan.</td>
<td>Disaggregated discipline &amp; achievement (achievement tests, DIBELS, SWIS) Action Plan/CSP Worksheet Decision rules for moving between the tiers of support</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>CRP</td>
<td>8. An analysis of culturally responsive practices was conducted as part of collaborative strategic planning.</td>
<td>No analysis of culturally responsive practices was conducted.</td>
<td>CRP section of the Team Survey or the entire CRP survey was completed but the results are not reflected in the building plan.</td>
<td>CRP section of the Team Survey or the entire CRP survey was completed and the results are reflected in the building plan.</td>
<td>Action Plan/CSP Worksheet Completed Team Worksheet Completed CRP Survey</td>
<td>0-2</td>
</tr>
<tr>
<td>CSP 5.1A</td>
<td>9. The leadership team insure there is a cycle of planning for continuous learning opportunities to insure high levels of implementation of the model.</td>
<td>No evidence of a plan for providing continuous learning opportunities.</td>
<td>Some evidence of a plan for providing continuous learning opportunities (e.g. a listing of learning topics/opportunities related to the ISM model).</td>
<td>Clear documentation of action steps for providing continuous learning opportunities for adults in the school community including family members and including specific content, context and process focused on raising the achievement of all students and closing achievement gaps of under performing student subgroups.</td>
<td>PD schedule PD Plan</td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>10. The leadership team has a plan for training new staff in the ISM model.</td>
<td>No written plan.</td>
<td>A plan is in place, but has not been implemented OR a plan was implemented but without a written agenda.</td>
<td>A plan with a written agenda is in place and has been implemented.</td>
<td>PD Plan PD agendas</td>
<td></td>
</tr>
<tr>
<td>SBR 5.3T A</td>
<td>11. Written components of instruction and intervention (instructional integrity checklist) exist at Tier 1.</td>
<td>Instruction and intervention are not written down.</td>
<td>Some but not all instruction is captured in a written script or instructional integrity checklist.</td>
<td>Detailed instruction or intervention plans are written down (scripted).</td>
<td>Written description of instruction and intervention Instructional integrity checklist</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>DBDM</td>
<td>12. For all schools with classrooms from Kg-6th grade there is a school-wide electronic (universal) data system in place for academic screening across all grades (DIBELS, AIMSWEB) and all students are screened at least 3 times per year.</td>
<td>No system in place in up to 49% of the classrooms.</td>
<td>System is in place and being used by a least 50% of classrooms.</td>
<td>100% of classrooms screen with data at least 3 times per year.</td>
<td>DIBELS &amp; AIMSWEB reports</td>
<td>0-2</td>
</tr>
<tr>
<td>DBDM</td>
<td>13. A school-wide electronic behavioral reporting system (SWIS or an equivalent) to monitor discipline referral data is in place and consistently implemented.</td>
<td>No system is in place.</td>
<td>System has been adopted and implementation has begun.</td>
<td>System is in place and entries are no more than 2 weeks behind.</td>
<td>SWIS reports or equivalent</td>
<td>0-2</td>
</tr>
<tr>
<td>3T</td>
<td>14. A small (3-5) set of positive behavioral expectations has been defined, and is posted school wide in appropriate places.</td>
<td>No expectations are defined or expectations are posted in less than 4 places.</td>
<td>Expectations are defined and posted in 5-7 of expected places.</td>
<td>Expectations are defined and posted in 8-10 of expected places.</td>
<td>(8 defined places, including a sample of classrooms observed)</td>
<td>0-2</td>
</tr>
<tr>
<td>3T</td>
<td>15. Lesson plans (or written guidance) to teach expectations have been developed, including developmentally appropriate examples for all school settings (all classrooms, and major common areas).</td>
<td>No lesson plans have been developed.</td>
<td>Lesson plans (or written guidance) are developed and include developmentally appropriate examples for either out of classroom or for classroom areas.</td>
<td>Lesson plans (or written guidance) are developed and include developmentally appropriate examples for both major out of classroom areas and classroom areas.</td>
<td>(A copy of lesson plans, or similar documents, with examples for classrooms and major out of classroom areas are examined)</td>
<td>0-2</td>
</tr>
<tr>
<td>3T</td>
<td>16. Teachers asked know 80% of the expectations.</td>
<td>0-50% of teachers state 80% of expectations.</td>
<td>51-89%</td>
<td>90-100%</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>3T</td>
<td>17. Students asked know 80% of the common expectations.</td>
<td>0-50% of students state 80% of expectations.</td>
<td>51-89%</td>
<td>90-100%</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>3T</td>
<td>18. Teachers state that expectations have been taught this year.</td>
<td>0-50% state that expectations have been taught.</td>
<td>51-89%</td>
<td>90% say yes</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>3T</td>
<td>19. A school-wide system to systematically provide positive consequences for meeting expectations is written and implemented.</td>
<td>No system is in place.</td>
<td>Either 75% of teachers state that they have given a consequence as specified in the system within the last week OR at least 75% of students state that they have received a consequence within the last month.</td>
<td>75% of teachers state that they have given a consequence as specified in the system within the last week AND at least 75% of students state that they have received a consequence within the last month.</td>
<td>(10 staff, and 10-15 students are selected randomly and interviewed)</td>
<td></td>
</tr>
</tbody>
</table>
| 3T       | 20. There is a written continuum of responses for dealing with specific behavior violation. | No                                     | A listing of possible responses exists but there is no clear written continuum. | Yes, a written continuum with clear steps exists. | Staff Handbook  
Discipline code – Code of conduct                                                                                                                             |       |
<p>| 35T      | 21. Do 90% of staff asked agree with the administration on what behavior problems are office managed and what problems are classroom managed. | 0-50%                                  | 51-89%                                 | 90-100%                               | Interviews                                                                                                                                                       |       |</p>
<table>
<thead>
<tr>
<th>Features</th>
<th>Evaluation Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Data Source</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T</td>
<td>22. A core reading program is in place at Tier 1 for all grades.</td>
<td>No evidence of a core curriculum.</td>
<td>Evidence of a core curriculum at some grade levels.</td>
<td>Evidence of a core curriculum at all grade levels.</td>
<td>Interviews</td>
<td>0-2</td>
</tr>
<tr>
<td>SBR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T</td>
<td>23. The core reading curriculum has been analyzed for overlap with Ohio's academic content standards.</td>
<td>No evidence of analysis.</td>
<td>Evidence of analysis at some grade levels.</td>
<td>Evidence of analysis at all grade levels served.</td>
<td>Curriculum Maps</td>
<td></td>
</tr>
<tr>
<td>SBR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T</td>
<td>24. The core reading program (Tier 1) and instructional programs or practices used at Tier 2 have been analyzed relative to research-based practices. Necessary changes or supplementation have been made.</td>
<td>No evidence of analysis of the core reading program and instructional programs or practices.</td>
<td>Evidence of analyzing the core reading program and Tier 2 instructional practices at some grade levels – OR – Evidence of analysis of the core reading program and Tier 2 instructional practices without any changes or supplementation made.</td>
<td>Clear documentation of analysis of the core reading program and documentation of any necessary changes or supplementation made.</td>
<td>Consumer’s Guide, PET, and/or Curriculum Map and review of Universal Screening Data (DIBELS/CBM) Research Literature</td>
<td></td>
</tr>
<tr>
<td>SBR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T</td>
<td>25. Targeted literacy supports are provided for students who need them.</td>
<td>No</td>
<td>Yes for some grade levels or some students.</td>
<td>Yes for all grade levels and all students.</td>
<td>Interview Teaching schedules</td>
<td></td>
</tr>
<tr>
<td>SBR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3T</td>
<td>26. Targeted behavior supports are provided for students who need them.</td>
<td>No</td>
<td>Yes for some grade levels or some students.</td>
<td>Yes for all grade levels and all students.</td>
<td>Interview Teaching schedules</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C:

IMPLEMENTATION EVALUATION TOOL PHASE III
### Key Features

#### Administrative Leadership
- **Evaluation Item:** There is a representative building leadership team.
  - **Score:**
    - **0:** The building leadership team lacks diverse representation.
    - **1:** The building leadership team has diverse representation but does not include parent membership.
    - **2:** The building leadership team has diverse representation (administration, general education, intervention specialists, etc.) including parent membership.
  - **Data Source:** Meeting notes with a list of members present

#### Collaborative Strategic Planning (CPS)
- **Evaluation Item:** The building leadership team meets regularly to guide the action planning process.
  - **Score:**
    - **0:** No evidence that the building leadership team meets regularly.
    - **1:** The building leadership team meets one-two times per year.
    - **2:** The building leadership team meets regularly (at least 3 times).
  - **Data Source:** Meeting schedule

#### Collaborative Strategic Planning (CPS)
- **Evaluation Item:** The leadership team has a plan for supporting (onboarding) new team members in learning the components of the model and implementation activities.
  - **Score:**
    - **0:** No evidence of a plan for supporting new team members in learning the components and activities of the model.
    - **1:** Some evidence of a plan for supporting new team members in learning the components and activities of the model reported or documented.
    - **2:** Clear documentation of action steps for supporting new team members in learning the components and activities of the model.
  - **Data Source:** Written documentation of the plan

#### Administrative Leadership
- **Evaluation Item:** The principal is a member of the building leadership team and regularly attends meetings.
  - **Score:**
    - **0:** The principal does not attend.
    - **1:** The principal occasionally attends meetings (at least 90% of meetings).
    - **2:** The principal regularly attends meetings (at least 90% of meetings).
  - **Data Source:** Meeting notes listing participants

#### Administrative Leadership
- **Evaluation Item:** The principal demonstrates understanding of the Key Features of ISM.
  - **Score:**
    - **0:** The principal does not understand the Key Features of ISM.
    - **1:** The principal can talk knowledgeably about some Key Features of ISM, but not all of them.
    - **2:** The principal speaks knowledgeably about the Key Features of ISM.
  - **Data Source:** Principal interview

---

<table>
<thead>
<tr>
<th>Features</th>
<th>Evaluation Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Data Source</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL CSP</td>
<td>1. There is a representative building leadership team.</td>
<td>The building leadership team lacks diverse representation.</td>
<td>The building leadership team has diverse representation but does not include parent membership.</td>
<td>The building leadership team has diverse representation (administration, general education, intervention specialists, etc.) including parent membership.</td>
<td>Meeting notes with a list of members present</td>
<td>0-2</td>
</tr>
<tr>
<td>CSP</td>
<td>2. The building leadership team meets regularly to guide the action planning process.</td>
<td>No evidence that the building leadership team meets regularly.</td>
<td>The building leadership team meets one-two times per year.</td>
<td>The building leadership team meets regularly (at least 3 times).</td>
<td>Meeting schedule</td>
<td>0-2</td>
</tr>
<tr>
<td>AL CSP</td>
<td>3. The leadership team has a plan for supporting (onboarding) new team members in learning the components of the model and implementation activities.</td>
<td>No evidence of a plan for supporting new team members in learning the components and activities of the model.</td>
<td>Some evidence of a plan for supporting new team members in learning the components and activities of the model reported or documented.</td>
<td>Clear documentation of action steps for supporting new team members in learning the components and activities of the model.</td>
<td>Written documentation of the plan</td>
<td>0-2</td>
</tr>
<tr>
<td>AL</td>
<td>4. The principal is a member of the building leadership team and regularly attends meetings.</td>
<td>The principal does not attend.</td>
<td>The principal occasionally attends meetings (at least 90% of meetings).</td>
<td>The principal regularly attends meetings (at least 90% of meetings).</td>
<td>Meeting notes listing participants</td>
<td>0-2</td>
</tr>
<tr>
<td>AL</td>
<td>5. The principal demonstrates understanding of the Key Features of ISM.</td>
<td>The principal does not understand the Key Features of ISM.</td>
<td>The principal can talk knowledgeably about some Key Features of ISM, but not all of them.</td>
<td>The principal speaks knowledgeably about the Key Features of ISM.</td>
<td>Principal interview</td>
<td>0-2</td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score 0-2</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>CSP</td>
<td>6. The building leadership team uses the Collaborative Strategic Planning Process to develop a building action plan that includes tiered systems of support.</td>
<td>No evidence of the Collaborative Strategic Planning process documented on action plan.</td>
<td>Some evidence of the Collaborative Strategic Planning process documented on action plan.</td>
<td>Clear documentation of each step of the Collaborative Strategic Planning process is documented on action plan.</td>
<td>CSP Worksheet/Action plan (must be based on school-wide data)</td>
<td>0-2</td>
</tr>
<tr>
<td>CRP</td>
<td>7. As part of collaborative strategic planning, data on suspension/expulsion and achievement (test scores, DIBELS, CBM, SWIS) are disaggregated by NCLB subgroups and used in decision-making.</td>
<td>Data are not disaggregated.</td>
<td>Either suspension/expulsion or achievement data analyzed but not both, OR data are disaggregated but there is no evidence that they were used in decision making (i.e. not reflected in the Action Plan).</td>
<td>Data are disaggregated, and the data are used for decision making in the Collaborative Strategic Planning process and are reflected on the Action Plan.</td>
<td>Disaggregated discipline &amp; achievement (achievement tests, DIBELS, SWIS)</td>
<td>0-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Action Plan/CSP Worksheet or Analysis Guides</td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>8. An analysis of culturally responsive practices was conducted as part of collaborative strategic planning.</td>
<td>No analysis of culturally responsive practices was conducted.</td>
<td>CRP section of the Team Survey or the entire CRP survey was completed but the results are not reflected in the building plan.</td>
<td>CRP section of the Team Survey or the entire CRP survey was completed and the results are reflected in the building plan.</td>
<td>Action Plan/CSP Worksheet or Analysis Guides</td>
<td>0-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Completed Team Survey</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Completed CRP Survey</td>
<td></td>
</tr>
<tr>
<td>CSP AL</td>
<td>9. The leadership team insures there is a cycle of planning for continuous learning opportunities to insure high levels of implementation of the model.</td>
<td>No evidence of a plan for providing continuous learning opportunities.</td>
<td>Some evidence of a plan for providing continuous learning opportunities (e.g. a listing of learning topics/opportunities related to the ISM model).</td>
<td>Clear documentation of action steps for providing continuous learning opportunities for adults in the school community including family members and including specific content, context and process focused on raising the achievement of all students and closing achievement gaps of under performing student subgroups.</td>
<td>PD schedule</td>
<td>0-2</td>
</tr>
<tr>
<td>CRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PD Plan</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-------</td>
</tr>
<tr>
<td>CSP AL</td>
<td>10. The leadership team has a plan for training new staff in the ISM model.</td>
<td>No written plan.</td>
<td>A plan is in place, but has not been implemented OR a plan was implemented but no written agenda exists.</td>
<td>A plan with a written agenda is in place and has been implemented.</td>
<td>PD Plan PD agendas</td>
<td>0-2</td>
</tr>
<tr>
<td>SBR 3T CRP AL</td>
<td>11. Written components of instruction and intervention (instructional integrity checklist) exist at Tier 1.</td>
<td>Instruction and intervention are not written down.</td>
<td>Some but not all instruction is captured in a written script or instruction integrity checklist.</td>
<td>Detailed instruction or intervention plans are written down (scripted).</td>
<td>Written description of instruction and intervention Instructional integrity checklist</td>
<td>0-2</td>
</tr>
<tr>
<td>SBR 3T CRP AL</td>
<td>12. Written components of instruction and intervention (instructional integrity checklist) are used to ensure criteria for accurate implementation are met and used by those who deliver the instruction at Tier 1.</td>
<td>Instruction and intervention not written down or criteria are not met.</td>
<td>Instructional integrity checklists are used in training staff who will implement the instruction.</td>
<td>Instructional integrity checklists are used to train staff to criterion prior to implementing the instruction.</td>
<td>Written description of instruction and intervention Instructional integrity checklist</td>
<td>0-2</td>
</tr>
<tr>
<td>3T SBR</td>
<td>13. An instructional integrity checklist is used to ensure a high level of implementation of the instruction and intervention at Tier 1.</td>
<td>Instruction and intervention not implemented as planned or no evidence of an instructional integrity checklist.</td>
<td>Instructional integrity checklists indicate some instruction and intervention is implemented as planned.</td>
<td>Instructional integrity checklists indicate instruction and intervention are implemented as planned.</td>
<td>Instructional integrity checklist</td>
<td>0-2</td>
</tr>
<tr>
<td>SBR 3T</td>
<td>14. Student interventions at Tier 2 include evidence of implementation.</td>
<td>No evidence of intervention implementation can be provided.</td>
<td>Intervention scripts are available for some interventions/students and/or implementation integrity data available for some interventions/students.</td>
<td>Intervention scripts with implementation integrity data that indicate compliance are available for students receiving interventions at Tier 2.</td>
<td>Scripts w/implementation integrity data</td>
<td>0-2</td>
</tr>
<tr>
<td>SBR 3T</td>
<td>15. Individual student interventions at Tier 3 are designed using research-based practices and include evidence of implementation.</td>
<td>Intervention not research-based and/or evidence of intervention implementation can be provided for less than 40% of cases.</td>
<td>Individual student intervention scripts are research-based and/or intervention scripts are available for 3 out of 5 cases and/or implementation integrity data available for 40-80% of cases.</td>
<td>Individual student intervention scripts are research-based with implementation integrity data that indicate compliance for more than 80% of student cases at Tier 3.</td>
<td>Scripts w/implementation integrity Research literature</td>
<td>0-2</td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------------</td>
<td>-------</td>
</tr>
</tbody>
</table>
| DBDM     | 16. A system is in place for using academic data to match students to services across the 3 tiers. | No documentation of a decision making system. | Decision rules described by staff, but not in writing, or across some, but not all, grade levels. | Decision rules for using data to match students to tiers of service are in writing and include:  
- A process for matching data to services  
- A process for using data to move between tiers  
- Agreement between staff and principal | Universal screening data (e.g. DIBELS, AIMSWEB, Excel, VIP)  
Interviews  
Team Survey  
Decision rules for moving between tiers of support | 0-2 |
| DBDM     | 17. A system is in place for using behavior data to match students to services across the 3 tiers. | No documentation of a decision making system. | Decision rules described by staff, but not in writing, or across some, but not all, grade levels. | Decision rules for using data to match students to tiers of service are in writing and include:  
- A process for matching data to services  
- A process for using data to move between tiers  
- Agreement between staff and principal | Universal screening data (e.g. SWIS)  
Interviews  
Team Survey and Analysis  
Decision rules for moving between tiers of support | 0-2 |
<p>| DBDM     | 18. For all schools with classrooms from Kg-6th grade there is a school-wide electronic (universal) data system in place for academic screening across all grades (DIBELS, AIMSWEB) and all students are screened at least 3 times per year. | No system in place in up to 49% of the classrooms. | System is in place and being used by a least 50% of classrooms. | 100% of classrooms screen with data at least 3 times per year. | DIBELS &amp; AIMSWEB reports |</p>
<table>
<thead>
<tr>
<th>Features</th>
<th>Evaluation Item</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Data Source</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBDM</td>
<td>19. A school-wide electronic behavioral reporting system (SWIS or an equivalent) to monitor discipline referral data is in place and consistently implemented.</td>
<td>No system in place.</td>
<td>System has been adopted and implementation has begun.</td>
<td>System is in place and entries are no more than 2 weeks behind.</td>
<td>SWIS reports or Equivalent Copy of discipline referral form</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>20. A small (3-5) set of positive behavioral expectations has been defined, and is posted school wide in appropriate places.</td>
<td>No expectations defined or posted in less than 4 places.</td>
<td>Expectations are defined and posted in 5-7 of expected places.</td>
<td>Expectations are defined and posted in 8-10 of expected places.</td>
<td>(8 defined places, including a sample of classroom are observed)</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>21. Lesson plans (or written guidance) to teach expectations have been developed, including developmentally appropriate examples for all school settings (all classrooms, and major common areas).</td>
<td>No lesson plans have been developed.</td>
<td>Lesson plans (or written guidance) are developed and include developmentally appropriate examples for either out of classroom or for classroom areas.</td>
<td>Lesson plans (or written guidance) are developed and include developmentally appropriate examples for both major out of classroom areas and classroom areas.</td>
<td>(A copy of lesson plans, or similar documents, with examples for classrooms and major out of classroom areas are examined)</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>22. Teachers asked know 80% of the expectations.</td>
<td>0-50% of teachers state 80% of expectations.</td>
<td>51-89%</td>
<td>90-100%</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>23. Students asked know 80% of the common expectations.</td>
<td>0-50% of students state 80% of expectations.</td>
<td>51-89%</td>
<td>90-100%</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>24. Teachers state that expectations have been taught this year.</td>
<td>0-50% state that expectations have been taught.</td>
<td>51-89%</td>
<td>90%</td>
<td>Interview</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>25. A school-wide system to systematically provide positive consequences for meeting expectations is written and implemented.</td>
<td>No system is in place.</td>
<td>Either 75% of teachers state that they have given a consequence as specified in the system within the last week OR at least 75% of students state that they have received a consequence within the last month.</td>
<td>75% of teachers state that they have given a consequence as specified in the system within the last week AND at least 75% of students state that they have received a consequence within the last month.</td>
<td>(10 staff, and 10-15 students are selected randomly and interviewed)</td>
<td>0-2</td>
</tr>
<tr>
<td>JT</td>
<td>26. There is a written continuum of responses for dealing with specific behavior violation.</td>
<td>No</td>
<td>A plan is in place, but has not been implemented OR a plan was implemented but without a written agenda.</td>
<td>A plan with a written agenda is in place and has been implemented.</td>
<td>PD Plan PD agendas</td>
<td>0-2</td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>Data Source</td>
<td>Score</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>----------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>JT</td>
<td>27. Do 90% of staff asked agree with the administration on what behavior problems are office managed and what problems are classroom managed.</td>
<td>0-50%</td>
<td>51-89%</td>
<td>90-100%</td>
<td>Interview</td>
<td></td>
</tr>
<tr>
<td>JT</td>
<td>28. A core reading program is in place at Tier 1 for all grades.</td>
<td>No evidence of a core curriculum.</td>
<td>Evidence of a core curriculum at some grade levels.</td>
<td>Evidence of a core curriculum at all grade levels.</td>
<td>Interviews</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JT</td>
<td>29. The core reading curriculum has been analyzed for overlap with Ohio’s academic content standards.</td>
<td>No evidence of analysis.</td>
<td>Evidence of analysis at some grade levels.</td>
<td>Evidence of analysis at all grade levels served.</td>
<td>Curriculum Maps</td>
<td></td>
</tr>
<tr>
<td>JT</td>
<td>30. The core reading program and instructional programs or practices used at Tiers 2 &amp; 3 have been analyzed relative to research-based practices. Necessary changes or supplementation have been made.</td>
<td>No evidence of analysis of the core reading program and instructional programs or practices.</td>
<td>Evidence of analyzing the core reading program and Tier 2 instructional practices at some grade levels – OR – Evidence of analysis of the core reading program and Tier 2 instructional practices without any changes or supplementation made.</td>
<td>Clear documentation of analysis of the core reading program and documentation of any necessary changes or supplementation made.</td>
<td>Consumer’s Guide, PET, and/or Curriculum Map and review of Universal Screening Data (DIBELS/CBM) Research Literature</td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>Evaluation Item</td>
<td>Score</td>
<td>Data Source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JT</td>
<td>31. Targeted literacy supports are provided for students who need them.</td>
<td>0</td>
<td>Interview, Teaching schedules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>Yes for some grade levels or some students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Yes for all grade levels and all students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JT</td>
<td>32. Targeted behavior supports are provided for students who need them.</td>
<td>0</td>
<td>Interview, Teaching schedules</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>Yes for some grade levels or some students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Yes for all grade levels and all students.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JT CSP</td>
<td>33. Team uses Collaborative Problem Solving to develop an intervention plan and inform decisions made at Tier 3.</td>
<td>0</td>
<td>Individual CPS Worksheet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evidence of Collaborative Problem Solving can be provided for less than 40% of the Tier 3 case files reviewed.</td>
<td>1</td>
<td>Some evidence of team use of Collaborative Problem Solving as documented on the Problem Solving Worksheet for 40-80% of the Tier 3 case files reviewed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Clear documentation of outcomes of each step of the Collaborative Problem Solving process on the Problem Solving Worksheet (based on data) in greater than 80% of the Tier 3 case files reviewed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBDM</td>
<td>34. Data are collected frequently and used to inform decisions made at Tier 2 and Tier 3.</td>
<td>0</td>
<td>Individual CPS Worksheet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evidence of frequent (i.e., weekly for Tier 3, every other week for Tier 2) data collection can be provided in less than 40% of the case files reviewed.</td>
<td>1</td>
<td>Some evidence of frequent (i.e., weekly for Tier 3, every other week for Tier 2) data collection and/or use of data to make decisions in 40-80% of case files reviewed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Evidence of weekly for Tier 3 and every other week for Tier 2 progress monitoring data collected AND graphed in greater than 80% of case files reviewed with evidence of use of data to make decisions.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D:

CONSENT LETTER FOR RELEASE OF DATA
Dear XXXXXX:

My name is Amity Noltemeyer and I am a doctoral candidate in the School Psychology program at Kent State University. For my doctoral dissertation, I would like to learn more about the relationship between Integrated Systems Model (ISM) implementation and student outcomes. I am particularly interested in identifying (a) the effect of level of ISM implementation on reading and behavior outcomes, and (b) the alterable factors within an ISM model that have the most influence on improving student outcomes.

Because your school has been involved in implementing ISM, I am writing to request permission to access two sources of information from your school to use for this study.

First, I am requesting permission to access evaluation data obtained through implementation of the State Improvement Grant (SIG) 2004-2007. Should consent be provided for the release of these data, they will be released to me by Karen R. Schaeffer, Ph.D., Regional SIG Coach. Specifically, the following information on your school for each year it has been collected since 2004 will be released: (a) Implementation Evaluation Tool (IET) scores by item, including any inter-rater reliability data; (b) DIBELS or AimsWeb data for each benchmark assessment disaggregated by school and grade-level; and (c) frequency of office discipline referrals (via SWIS reports if available).

Second, I am requesting for your school to provide me with student-level DIBELS or AimsWeb data with student names removed or blinded. Specifically, I am seeking individual student scores for each benchmark assessment administered from the 2004-2005 to 2007-2008 school years for all grade-levels in which data was collected at your school. The use of student-level data is an important component of the study because it will allow more sophisticated analysis techniques (e.g., hierarchical linear modeling), resulting in a more precise estimation of the effects of ISM on student outcomes. Should you provide consent for the release of these data, they could be released to me in a manner most convenient for you (e.g., exporting the data to Excel without the names
using the DIBELS data farming technique. Even partial release of these data is requested (e.g., if you only have data from one year or a few grade levels).

All released information will be treated with high confidentiality standards. After obtaining the data, all school buildings and districts will be assigned pseudonyms and a pseudonym “key” will be kept in a locked file cabinet. All data will be entered into statistical software and stored on a flashdrive in a locked file cabinet. Data will be deleted from the flashdrive after five years. Any printouts of the data will also be stored in the locked file cabinet and shredded after five years. No building or district names will be revealed in any subsequent dissemination of the data (i.e. conference presentations, manuscripts/chapters for publication). However, results of the analyses of the SIG data may be shared with the regional SIG Coach.

It is anticipated the aforementioned data will be analyzed using univariate statistics, multivariate statistics, and/or multilevel modeling (e.g., hierarchical linear modeling). The data will not be analyzed or disseminated without Kent State University Institutional Review Board approval. Upon completion of the dissertation, the results will be disseminated via professional writing, presentations, and conference submissions. Again, no district or building names will be used in these presentations and publications.

I appreciate your willingness to consider this request, and I am confident that the results of this project can be used to inform future practice to improve outcomes for students. Your school has already done so much to improve students’ academic and behavioral success, and I feel fortunate to have been provided the opportunity to contact you. Please sign and return the attached form in the enclosed stamped envelope to provide consent for the release of one or both data sets. Please do not hesitate to contact me if I can provide you with additional information or clarification. My primary advisor, Dr. Frank Sansosti, is also available to address questions or concerns (330-672-0059, fsansost@kent.edu).

Sincerely,

Amity Noltemeyer, Ed.S., NCSP
Doctoral Student in School Psychology
anolteme@kent.edu
Integrated Systems Model Implementation and Student Outcomes

CONSENT STATEMENT:

I fully understand the previously described information and give consent for the release of the following data to be used for Amity Noltemeyer’s doctoral dissertation:

(Check all that apply. Both sources of data are preferred.)

________ I give the Regional SIG Coach permission to release the aforementioned ISM evaluation data for ___________________________ school

________ I agree to release the requested student-level DIBELS or AimsWeb data, with names withheld, to Amity Noltemeyer for use for her doctoral dissertation (see the third paragraph of the letter above for more information).

(Note: If only part of these data will be released, please specify here _____________________________)

_________________________________________  _____________________________
Name (print)                                  Signature

_________________________________________  _____________________________
Title                                          Date
Integrated Systems Model Implementation and Student Outcomes

CONSENT STATEMENT:

I fully understand the previously described information and give consent for the release of the following data to be used for Amity Noltemeyer’s doctoral dissertation:

(Check all that apply. Both sources of data are preferred.)

[ ] I give the Regional SIG Coach permission to release the aforementioned ISM evaluation data for ________________ school

[ ] I agree to release the requested student-level DIBELS or AimsWeb data, with names withheld, to Amity Noltemeyer for use for her doctoral dissertation (see the third paragraph of the letter above for more information).

(Note: If only part of these data will be released, please specify here ______________________________________________________________________)

Name (print) ___________________________ Signature ___________________________

Title ___________________________ Date ___________________________
REFERENCES
REFERENCES


Berends, M., Bodilly, S., & Kirby, S. (2002). Facing the challenge of whole-school reform. Santa Monica, CA: RAND.


behavior analysis with children. *Journal of Applied Behavior Analysis, 26*, 257-263.


