A Psychometric Examination of the School Implementation Scale

By

Kayla Supon Carter

Submitted to the graduate degree program in Educational Psychology and Research in and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____
Dr. John Poggio, Chair

____
Dr. Amy Gaumer Erickson

____
Dr. David Hansen

____
Dr. Suzanne Rice

____
Dr. Argun Saatcioglu

____
Dr. William Skorupski

Date Defended: September 15, 2017

The Dissertation Committee for Kayla Supon Carter

certifies that this the approved version of the following dissertation:

A Psychometric Examination of

the School Implementation Scale

____

Chair: Dr. John Poggio

Date Approved: October 18, 2017

Abstract

The purpose of this study is to further enhance the existing reliability and validity evidence related to the School Implementation Scale (SIS) and its use for evaluating the implementation of different integrated academic and behavior tiered Response-To-Intervention (RTI) models. Previous methodology conducted for validation of the instrument has been expanded to examine the factor structure of the measurement instrument and the relationship between those factors across two overall state-implemented educational initiatives and across two different types of integrated academic and behavior multi-tiered systems of support within and between two states. The exploratory and confirmatory factor analyses that have been conducted in both SPSS and Mplus on the five datasets (Midwestern Collaborative Work, Western Effective Behavioral and Instructional Support Systems, Midwestern School-Wide Positive Behavior Supports, Western Schoolwide Positive Behavioral and Instructional Supports or Response-To-Intervention, and both the Midwestern and Western states combined) consisting of the common 28 School Implementation Scale items reveals that the 4-factor structure specified for the SIS does not adequately fit any of the datasets. The SIS, however, initially consisted of 31 items. Statistical analysis of the 2014 Midwestern state's dataset with 31 items also reveals that the data does not adequately fit the 4-factor structure specified for the SIS. Therefore, the 4-factor structure may have been an incorrectly specified model for the School Implementation Scale, and should be reviewed then revised. This study concludes that the SIS with its current 4-factor structure, for either 28 or 31 items, is not a valid and reliable measure of the implementation level of integrated academic and behavioral multi-tiered systems, and should not be used across various state populations and different forms of multi-tiered support systems.

Acknowledgements

I dedicate this doctoral dissertation to my family, friends, co-workers, and committee members...

I need to thank my wonderful husband, Ted, and my intelligent and talented children, Bailey and Alex, for their patience and understanding and encouragement- even when I had to physically (or mentally) miss out on other important events or occasions. I hope “Dr. Mom” was worth it! I also need to express my gratitude to my Mom, Dad, and Brother for their unconditional love and support in all of the assorted endeavors throughout my life.

There have been several very special friends that have helped me since the beginning of this journey, and continuously cheered me on throughout the entire process. Among these individuals are fellow graduate students from the Educational Psychology department, various co-workers during my graduate research assistant career, my current co-workers at Washburn University, and a few other vitally important people without whom this quest could not have been completed.

“To Sir, With Love”- Dr. John Poggio, I appreciate your never-ending patience and encouragement during my extended timeline in completing this dissertation, and for not allowing me to become a person with the all-to-popular “A.B.D.” designation. I have learned so much from you regarding program evaluation, teaching, and persistence- I will never forget your wonderful mentoring and will hear your voice as part of my conscience nearly every day.

A special thanks to Dr. Amy Gaumer Erickson at Research Collaboration for the use of the School Implementation Scale and data for this project. I also wish to express my gratitude to my other committee members- Dr. David Hansen, Dr. Suzanne Rice, Dr. Argun Saatcioglu, and Dr. William Skorupski- for their willingness to serve on my committee, for their constructive feedback, and for their determination to help me succeed.

Table of Contents

List of Tables

CHAPTER 1

Introduction

Gaumer Erickson, Noonan, and Jensen (2012) have developed a cost-effective, evidence-based treatment integrity measure to better understand the implementation of integrated academic and behavior tiered-structure models. The School Implementation Scale (SIS) has been designed for all school staff to complete within 5-10 minutes. Results are intended to be used in school improvement efforts, and to assist evaluators in quantifying the implementation fidelity of integrated academic and behavioral tiered Response-to-Intervention (RTI) models. The instrument (see Appendix A) consists of 31 Likert-type items in which respondents rate statements from 1 (Not True of Me) to 5 (Very True of Me), with the mid-point being labelled as Somewhat True of Me. The SIS encompasses four essential elements of successful schools: School Culture (9 items), Evidence-Based Practices (13 items), Ongoing Professional Development (4 items), and Family Involvement (5 items). The original Gaumer Erickson, et al. (2012) study evaluated the salient empirical properties and outcomes of the SIS within one integrated academic and behavior tiered RTI model. The analyses consisted of reliability estimates in the form of Cronbach's coefficient alphas on the internal consistency of the items, Exploratory Factor Analysis (EFA) to determine their factor structure of the items, One-Way Analyses of Variance (ANOVA) and *t*-tests to compare various group means, and Pearson Product-Moment Correlations of SIS results with student outcomes. The instrument showed acceptable validity and reliability, and moderate correlations were found between SIS scores and reading/writing achievement of students with disabilities (Gaumer Erickson, et al., 2012).

As the use of the SIS is increasing to include different integrated academic and behavior tiered RTI models in more states and school districts, additional statistical analyses should be conducted to examine more extensively the validity, reliability, and measurement invariance of the instrument and its resulting scores. The more advanced statistical methodology will facilitate a compilation of enhanced validity and reliability information regarding the instrument. The results of these analyses would assist in guiding the expanding use of the SIS, and help to ensure its appropriate employment and interpretation within various contexts. The additional reliability and validity evidence resulting from this research study related to the SIS will add to the results of the previous study, and will help evaluators and educators more knowledgably and accurately measure the implementation of integrated academic and behavior RTI models of tiered supports using the essential elements of successful schools. The supplementary psychometric information resulting from this study regarding the standardized SIS instrument could potentially justify the continued administration of the SIS for evaluation research, and lead to its valuable use in comparing various integrated academic and behavior tiered RTI models in terms of measuring and monitoring implementation. Effective strategies for monitoring implementation of integrated RTI models, those that meet *The Standards for Educational and Psychological Testing* (2014), should be shared across initiatives so as to advance the field of school-wide educational reform and evaluation. If the measurement instrument does not meet the standards, then educators and evaluators should be cautioned against making test score interpretations that are not supported.

"Measurement instruments that are collections of items combined into a composite score, and intended to reveal levels of theoretical variables not readily observable by direct means, are often referred to as a scale" (DeVellis, 2003, p. 8-9). Scales are created or employed when the phenomena to be measured exist because of a theoretical understanding, but these variables cannot be assessed directly. The measurement instrument must reflect the variables or constructs it is meant to assess, leading to examination of the relationships among proxies (such as scales) that are intended to represent the variable or construct of interest. Valid and reliable measurement instruments are imperative for quality educational research and evaluation. “A test is valid for measuring an attribute if: (a) the attribute exists, and (b) variation in the attribute causally produces variation in the measurement outcomes” (Borsboom, Mellenbergh, & van Heerden, 2004, p. 1061). A valid test or measurement instrument can convey the effect of variation in the attribute one intends to measure if variation in the attribute causes variation in the test scores, but “invalid or unreliable measures can harm a study to the same extent as poor study design or inadequate sample size” (Kimberlin & Winterstein, 2008, p. 2283). The foundation for evaluating the quality of a test or measurement instrument is provided by *The Standards for Educational and Psychological Testing* (2014), which also assists in evaluating the instrument’s effects or consequences of its use. These testing standards promote the “sound and ethical use of tests” (*Standards for Educational and Psychological Testing*, 1999, p.1), and “provide a frame of reference to assure that relevant issues are addressed” (p. 2) when creating or selecting a measurement instrument for research or program evaluation purposes.

“Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to the fairness for all examinees” (*Standards for Educational and Psychological Testing*, 1999, p.17). So becomes the charge for this investigation.

Statement of the Problem

Evaluators and school teams need access to measures of implementation for educational tiered support systems that center on the essential elements of effective school systems.

Evaluation measures can inform data-based decision-making within schools, evaluate outcomes of multiple integrated academic and behavior tiered response-to-intervention (RTI) models, and lead to the comparison of these tiered support models to provide a comprehensive, research-based framework for continued improvement in schools. Measures must be integrated to ensure they have the potential to inform school-wide data-based decision-making, to evaluate the implementation and effectiveness of the RTI model, and to compare implementation components across different general tiered RTI models such as School-Wide Positive Behavior Supports (SW-PBS) or Multi-Tier System of Supports (MTSS), and Response-To-Intervention (RTI). Various measurement instruments have been developed for these purposes, but not all measurement instruments are created equal. All scales, especially those used in educational evaluation, should be developed according to, and examined extensively following the guidance of *The Standards for Psychological and Educational Testing* (2014), *The Program Evaluation Standards* (2011), and the *AEA Guiding Principles for Evaluators* (2004). However, many evaluators are not familiar with the professional standards' specific recommendations and pitfalls to avoid when generating items and creating scales. Therefore, not all measurement instruments meet the professional standards for program evaluation or educational and psychological testing. "Adequate measures are a necessary condition for valid research," and "poor measurement imposes an absolute limit on the validity of the conclusions one can reach" (DeVellis, 2003, p.12). This study will examine important psychometric properties- the validity, reliability, and measurement invariance- of the School Implementation Scale (SIS) as a measurement instrument for use in evaluative studies of different integrated academic and behavior tiered RTI models through more advanced statistical and psychometric analyses than previously conducted.

Gaumer Erickson, et al. (2012) suggested that future research regarding the measure should include administration of the School Implementation Scale (SIS) in schools implementing various integrated academic and behavior tiered RTI models and continued administration of the SIS across multiple years to build the research base of measurement results. Data has been collected for two consecutive years, 2012-2013 and 2013-2014, from two different states, one Midwestern and one Western. These states implemented different overall educational initiatives and different integrated academic and behavior tiered RTI models. The Midwestern state's overall educational initiative is Collaborative Work (CW), which implements School-Wide Positive Behavior Supports (SW-PBS) as their integrated RTI model. The Western state's overall educational initiative was the Effective Behavioral and Instructional Support Systems (EBISS), and tiered models implemented include either Schoolwide Positive Behavioral and Instructional Supports (SWPBIS) or Response-To-Intervention (RTI). With this data that includes various integrated academic and behavior tiered RTI models, these research needs can be addressed. By expanding the previous validation study methodology through the use of more advanced statistical analysis than previously conducted, the existing reliability and validity evidence related to the SIS can be further enhanced.

The purpose of this study is to enhance and complement the existing reliability and validity evidence related to the School Implementation Scale and its use for evaluating different integrated academic and behavior tiered RTI models. This will be accomplished using the 2014 data by examining the factor structure of the measurement instrument and the relationship between those factors; testing the measurement invariance of the SIS constructs across two overall state-implemented educational initiatives (Midwestern CW and Western EBISS) and across the major types of integrated academic and behavior tiered RTI models within and between the two states (Midwestern SW-PBS/Western SWPBIS and Western RTI). These findings will be subject to cross-validation procedures using the 2013 data. The differences in the 2014 mean SIS results between various respondent groups by state and specific tiered RTI model being implemented will be examined to the extent that measurement invariance holds for the data.

Hypotheses

It is hypothesized that (1) the theorized 4-factor structure by Gaumer Erickson, et al. (2012) of the School Implementation Scale will fit the 2014 data, (2) that measurement invariance will hold for both the Midwestern and Western states' overall implemented educational initiatives, and (3) the types of integrated academic and behavioral tiered Response-to-Intervention models within and between the two states. Further, the School Implementation Scale overall, as well as its four essential elements of effective schools domains (School Culture, Evidence-Based Practices, Family Involvement, and On-going Professional Development), will provide reliable data across the state samples and types of integrated academic and behavior tiered response-to-intervention models for all school staff in terms of respondent's role, years of experience, and involvement with the leadership team. However, it is expected that differences in respondents' mean SIS and mean essential element domain scores result in terms of staff role (administrators greater than teachers greater than other staff), years at the school (smaller for the least and most experienced staff), and involvement with the leadership team (greater for team members). It is also expected that statistical interactions as mediators exist between the respondent demographic variables of staff role, years at the school, and involvement with the leadership team resulting in different patterns of SIS responses.

Research Questions

The Program Evaluation Standards (2011), the *AEA Guiding Principles for Evaluators* (2004), and *The Standards for Psychological and Educational Testing* (2014) that have been found to relate most directly and seem to be most critical in the accumulation of validity and reliability evidence for the School Implementation Scale assisted in formulating the following research questions to guide the current study.

1) For each state's 2014 data, and for each of the different integrated academic and behavior tiered RTI models, how well does the four-factor essential elements of successful schools structure from the School Implementation Scale fit the data; what are the relationships among the factors of the scale?
2) For the 2014 data (cross validated with 2013 data), to what extent does measurement invariance for the latent constructs measured by the School Implementation Scale, its four essential element domains, hold over states with different overall integrated academic and behavior tiered response-to-intervention models; specifically Collaborative Work (CW) from the Midwestern state and Effective Behavioral and Instructional Support Systems (EBISS) from the Western state?
3) For the 2014 data (cross validated with 2013 data), to what extent does measurement invariance for the latent constructs measured by the School Implementation Scale, its four essential element domains, hold over different major types of integrated academic and behavior tiered response-to-intervention models within and between states; specifically Response-to-Intervention (RTI) from the Western state and School-Wide Positive Behavior Supports (SW-PBS) from the Midwestern state and School-Wide Positive Behavior Interventions and Supports (SWPBIS) from the Western state?

Assuming measurement invariance of the School Implementation Scale holds to some level, research questions numbers 4 and 5 will only be conducted to the extent possible based on the results of measurement invariance testing.

4) For the 2014 data, does the School Implementation Scale instrument overall and its four essential elements of effective schools domains provide reliable data across the state samples (Midwestern CW and Western EBISS) and across the different major types of integrated academic and behavior tiered response-to-intervention models (Midwestern SW-PBS/ Western SWPBIS, and Western RTI) for all school staff in terms of respondent's role, years of experience, and involvement with the leadership team?
5) For the 2014 School Implementation Scale results, what are the differences between and patterns observed in respondents' mean SIS and mean essential element domain results in terms of staff role, years at the school, involvement with the leadership team, and the interactions of these variables for each state overall (Midwestern CW and Western EBISS) and for each states' major types of integrated academic and behavior tiered response-to-intervention models (Midwestern SW-PBS and Western SWPBIS, and Western RTI)?

Significance of Study

The additional reliability and validity evidence resulting from this research study of the School Implementation Scale will add valuable information to the results of the previous study by Gaumer Erickson, et al. (2012), and will help evaluators and educators more knowledgably and accurately measure the implementation of integrated academic and behavior RTI models of tiered supports using the essential elements of successful schools. The supplementary psychometric information regarding the standardized SIS instrument could potentially justify the continued administration of the SIS for evaluation research, and lead to its valuable use in comparing between various integrated academic and behavior tiered RTI models in terms of measuring and monitoring implementation; or results could indicate that the SIS necessitates changes before the instrument can be utilized across states or with different RTI models.

CHAPTER 2

Literature Review

Integrated Academic and Behavior Tiered Response-To-Intervention Models

Multi-Tiered Systems of Support (MTSS) is a process for school system reform which "extends the use of problem solving and data-driven decision making to include instructional strategies, classroom management, curriculum design, and professional development" (Dulaney, Hallam, & Wall, 2013, p.35). The MTSS framework is aimed at systematically supporting struggling students with behavioral and/or academic concerns, and provides three tiers of intervention with increasing intensity for students who are not responsive to the instruction and strategies at the lower tiers (Samuels, 2016). The collection and analysis of data to examine the extent of student improvement is an important component of any multi-tiered framework for progress monitoring, and to produce diagnostic information for guiding instruction and program placement decisions (Fuchs & Fuchs, 2006). MTSS consists of a set of evidence-based practices focused on student and educator support, quality service delivery across the system, and collaborative practices (Dulaney, et al., 2013). Response-to-Intervention and Positive Behavioral Interventions and Supports, also known as Positive Behavior Supports, are both examples of multi-tiered systems of supports (Samuels, 2016).

"Response to intervention (RTI) is widely used as a framework for providing high quality instruction and intervention that are matched to students' needs" (Mellard, Stern & Woods, 2011, p.1). Common features of RTI models include universal screening tools that allow teachers to determine which students need assistance, evidence-based interventions, multiple tiers of intensity so students who need more help receive a higher degree of intervention, and progress monitoring so educators have the data to know how well students are responding and make changes if necessary (Samuels, 2016). The basic RTI model is represented by a triangle as a conceptual blueprint for educators, showing that most students receive core instruction while progressively smaller groups require more intensive support. Tier 1 is the base of the triangle where all students receive high-quality evidence-based instruction. Moving up the triangle is Tier 2 where fewer students need more intensive instruction, and the top of the triangle is Tier 3 where even fewer students receive the most intensive services or are referred to be evaluated for special education (Samuels, 2016). However, as these multi-tiered models have evolved the conceptual representations have become more elaborate and complex to depict a collaborative system involving all forms of student support services.

RTI provides a comprehensive schema in which multi-tiered supports are utilized to prevent academic and behavioral difficulties, as well as to address existing academic and behavioral difficulties. When first introduced, RTI models focused on reading achievement and identification of students with learning disabilities while positive behavior support models focused on behavioral expectations, but research has acknowledged the inter-relatedness of academic and behavioral performance (Algozzine, Wang & Violette, 2011; Kalberg, Lane & Menzies, 2010; Lane, Kalhberg & Menzies, 2009). These integrated models include academic and behavioral interventions while continuing the overarching definition of RTI (Mellard et al, 2011). While RTI models have expanded to include integrated academic/behavior RTI models, evaluation of effectiveness for these models have remained focused on student level data (Sugai & Horner, 2009). Additional measures of treatment integrity within multi-tiered models have emerged to facilitate understanding of implementation and of essential elements for effective school systems that produce student-level academic and behavioral success (Sugai & Horner, 2009). Implementation fidelity is a critical component to producing the behavioral and academic improvements that can be potentially experienced under these educational initiatives (Samuels, 2016). The general category of Response-to-Intervention (RTI) includes information on system change models which have been shown to be effective in improving student achievement, such as School-Wide Positive Behavior Supports (SW-PBS).

According to the Positive Behavioral Interventions and Supports website (https://www.pbis.org/school), Schoolwide Positive Behavioral Interventions and Supports (SWPBIS) or SW-PBS is a decision-making framework that guides selection, integration, and implementation of evidence-based academic and behavioral practices for improving outcomes for all students. Four integrated elements are emphasized: 1) data for decision making, 2) measurable outcomes supported and evaluated by data, 3) practices with evidence of achievable outcomes, and 4) systems that efficiently and effectively support implementation of these practices ((https://www.pbis.org/school). SW-PBS is a process for creating safer and more effective schools by structuring the learning environment to support the academic and social success of all students (Horner, 2000). It applies behavioral analysis to the social problems created by inappropriate behaviors through a "committed focus on fixing environments, not people" (Horner, 2000, p.97). The process supports the adoption and long-term implementation of efficient and effective discipline throughout the school environment. SW-PBS methods are research-based, proven to significantly reduce the occurrence of problem behaviors in schools, and are supported by a three-tiered model (Simonsen, Sugai, & Negron, 2008). The three-tiered prevention logic model requires that all students receive supports at the universal or primary tier, more intensive behavioral supports be provided for some students in groups at the secondary tier, and highly individualized plans of support be provided at the intensive or tertiary tier (https://www.pbis.org/school). The continuum on which students experience supports is based on their responsiveness to intervention.

"Many of the practices and systems of SW-PBS are similar in nature to those of RTI; for example, universal screening, continuum of scientifically based behavioral interventions, data-based and team-driven decision-making structures, intervention integrity measures, and direct student performance measures" (Sugai & Horner, 2009, p.234). Both RTI and PBIS offer a range of interventions that are systematically applied to students based on their demonstrated level of need. A preventative approach to teaching academic and social behavior is a major component of RTI and of PBIS; as is the prerequisite that all academic and behavior interventions must be carried out with fidelity (Sandomierski, Kincaid, & Algozzine, 2007). Both RTI and SW-PBS or SWPBIS are "guided by an integration of data-based decision making, measurable outcomes, evidence-based practices, and systems for accurate and sustained implementation" (Sugai & Horner, 2009, p.234).

Although an important aspect of SWPBIS and RTI is the ongoing evaluation of program implementation fidelity (Bradshaw, Debnam, Koth, & Leaf, 2009), quality evaluation measures that consider the extent to which the models are implemented as intended across the whole school environment seem to be rare. "Self-report measures and external assessments of program fidelity have been created to determine the degree to which schools are implementing the key aspects" [of multi-tiered supports] (Bradshaw, et al., 2009, p.146), but most of these instruments consist of process checklists measuring the percentage of the model core features that are in place (Horner, Sugai & Lewis-Palmer, 2001), and typically utilize a yes/no or rating scale of basic components. Additionally, these checklists are usually only given to school implementation team members or administrators and coaches (Gaumer Erickson, et al., 2012), and the perceptions from all school staff are not considered in evaluating school-wide implementation of integrated academic and behavioral multi-tiered systems of student support.

Numerous measures have been developed to support School-Wide Positive Behavior Supports (SW-PBS) such as the Self-Assessment Tool, a 28-item observation and interview instrument created by Sugai, Lewis-Palmer, Todd, & Horner in 1999; and the Effective Behavior Support Survey, a tool with which all school staff rate the current status and priorities for improvement across behavior support systems developed by Todd, Lewis-Palmer, Horner, Sugai, Sampson, & Phillips in 2003. Another assessment of implementation or treatment fidelity related to SW-PBS includes the Benchmarks of Quality developed by Kincaid, Childs & George in 2005 which lists 53 behavioral strategy items to be rated as 'in place,' 'needs improvement,' or 'not in place,' and consensus on the overall level of implementation is reached through discussion (Cohen, Kincaid, & Childs, 2007). The Planning and Evaluation Tool for Effective Reading Supports is an instrument created by Kame'enui & Simmons in 2000 and is focused on goals, materials, time allocations, and other factors to rate overall implementation of school's K-3 reading program as 'in place,' 'partially in place,' or 'not in place.'

The Phases of Implementation Rubric is yet another measure of SW-PBS fidelity which was designed by the Vermont Department of Education in 2011. It is completed by the behavior coach or an observer who identifies implementation levels as emerging, implementing, or sustaining. Implementation rubrics provide a data source for school leaders to reflect on school-wide implementation of tiered models and plan for professional development (Liu, Alonzo & Tindal, 2011). The purposes of an implementation rubric are to outline operational definitions for implementation, provide illustrations of best practices, guide reflection on instructional practices and school improvement, and understand what treatment components were not implemented in schools that failed to see growth in student academic or behavioral achievement (Bradshaw et al., 2009).

Previous School Implementation Scale Research

Gaumer Erickson, et al. (2012) identified gaps in the existing fidelity measures being employed for integrated academic and behavior RTI models, stating that

"there are no existing measures that evaluate the implementation of the core features of integrated models from a whole school perspective in a cost-effective, minimally intrusive manner. Furthermore, measures that do exist are intervention specific and not appropriate for integrated models. Without treatment integrity data there is no way of understanding the variance in school gains" (p.36). They also claim that the data produced by existing treatment fidelity or implementation measures usually lack utilization focus, meaning that the data are not visually represented in meaningful and easily understandable ways for continual planning and improvement purposes (Gaumer Erickson, et al., 2012). Available measures of implementation fidelity have also focused primarily on the perceptions of small groups of individuals who constitute the leadership or implementation team. "RTI begins with the implementation of scientifically-based, schoolwide instructional interventions and promotes intervention at the first indication of nonresponse to traditional classroom instruction" (Bradley, Danielson, & Doolittle, 2007, p.8). Walker (2006), however, found that implementation ratings of the school leadership team members were significantly higher than those ratings of other school staff.

Gaumer Erickson, et al. (2012) developed the School Implementation Scale (SIS) to encompass the evidence-based essential elements of effective school systems designed and to be integrated into the school climate to drive decision making, facilitate innovation, and support students- school culture, ongoing professional development, evidence-based practices, and family engagement. The measure was developed and tested through iterative design processes which included: (a) framework conceptualization, (b) item development, (c) pilot testing, (d) item refinement, (e) additional pilot testing, (f) more item refinement, and (e) full implementation. An evaluation team of experts for the project helped identify criteria for development of the SIS. This team included university evaluators, state-level administrators, regional professional development providers, implementation coaches, and local-level educators. The team decided that the SIS needed: (1) to be cost-effective and efficient to administer, (2) obtain perceptions from all school staff not just those on leadership teams, and (3) provide results to schools in a format the was easy to interpret in order to inform team-level data based decision making and action planning. Thus, an online survey was created and data analyzed to produce school-level summary reports.

The framework for the various components assessed by the SIS are based on a comprehensive review of the essential elements of effective school systems as identified by National Center on Response to Intervention, the Technical Assistance Center on Positive Behavioral Interventions and Supports, Professional Learning Communities, Southern Regional Education Board High Schools that Work, and the US Dept. of Ed Reading First Initiative (Jenson, 2008). Item development for the SIS began with a literature review of the available school staff surveys within the frameworks of multi-tiered models. The Standards Assessment Inventory developed by National Staff Development Council was an important resource as it addressed some components of multi-tiered models, especially professional development, and was intended for all staff to complete. Additional items were developed to focus on individual implementation as opposed to perceived school-level implementation in order to address the inflation in scores of self-report data (Welkenhuysen-Gybels, Billiet & Cambre 2003). These items required the identification of personal practices, and because all staff complete the survey, composite results then provided whole-school implementation data. School staff were trained on the essential elements of effective school systems, which has been shown to increase the accuracy of self-assessment results (Irvine, 1983).

A version of the SIS which included 99 items developed on 5-pt Likert scale of 1 (not true of me) to 5 (very true of me) was reviewed by experts in integrated academic and behavior RTI models, and revised based on their feedback. In the spring of 2009 this 99-item survey was administered to all staff at 6 schools implementing the model, summary reports were developed and provided to the schools, and each school's action plan was reviewed to identify how survey results influenced needs and strategy identification. The implementation coach for each school also asked: (1) Was the data beneficial to the school leadership team? (2) Did the data assist the team in completing the action plan? (3) What ideas did the leadership team identify for improving the survey?

Reliability and exploratory factor analyses were conducted on the 99-item survey, and multiple items were removed because the number of items were considered too time-intensive for measurement of continued school-wide implementation. The team with expertise implementing tiered reform models (state, regional, and local professional development providers) reviewed the survey and participated in multiple focus groups discussing further revision of the instrument during the fall of 2009. This resulted in a 33-item survey that was piloted in the spring of 2010 in the 6 schools from the first pilot study, and an additional 8 schools implementing the same integrated multi-tiered model. Another round of reliability and exploratory factor analyses informed further item revision for the SIS.

During the 2010-2011 school year, 346 participants from 11 elementary and 3 middle schools (all public) completed the SIS. All of these school buildings were in the second year of implementing an integrated multi-tiered support model. According to the 2011 National Center for Statistics urbanicity classifications the schools consisted of two city, two suburban, three town, and six rural districts in all geographic regions of a Midwestern state. The schools had Free/Reduced lunch rates ranging from 21% to 71% of their populations, and the percentage of students from ethnically diverse backgrounds ranged from 1% to 95%. The number of staff per school ranged from 9 to 54; and included 294 teachers, 15 administrators, 24 other certified staff, and 11 non-certified staff.

Additionally, data other than the SIS were collected from 7 schools all in the first year of implementation (3 elementary, 1 middle, 3 high schools) within the original 14 districts for yearly and grade level comparisons. Three urban charter schools (1 elementary, 1 middle, and 1 K-8 school) from another state implementing a different integrated academic/behavior RTI model also completed the SIS during the 2011-12 school year, and data from these schools was used to analyze the reliability and validity of the SIS for use with other integrated RTI model(s).

The SIS was administered through a standardized process which included providing instructions to each school's principal with sample text for email distribution to staff. The principal then sent the information and survey link to all school staff and requested completion within 2 weeks. Staff accessed and completed the SIS (33 Likert-scale items and 3 demographic items: school name, professional role, and membership on the school leadership team), which required approximately 5-10 minutes. School summary reports provided a table displaying scores and means for each item and domain, as well as stacked bar graphs for the same data.

The statistical analyses conducted by Gaumer Erickson, et al. (2012) provided a fairly comprehensive description of validity and reliability of the SIS and initial outcomes related to implementation of RTI models. The analyses consisted of reliability estimates in the form of Cronbach's coefficient alphas to estimate the internal consistency of the items, Exploratory Factor Analysis to determine the factor structure of the items, One-Way Analyses of Variance (ANOVA) and T-tests to compare various group means, and Pearson Product-Moment Correlations of SIS results with student outcomes.

The reliability estimates resulting from the previous validity study for the School Implementation Scale are shown in Table 1. The overall coefficient alpha rendered for the 33-items SIS was 0.961. The essential element domain reliabilities equaled 0.808 for Family Engagement, 0.866 for Ongoing Professional Development, 0.888 for School Culture, and 0.902 for Evidence-Based Practices.

Table 1. Reliability Coefficient Estimates- Overall SIS, Domains, and Levels

School Implementation Scale	Number Items	Cronbach's Alpha
Overall SIS	33	0.961
Essential Element Domains	Number Items	Cronbach's Alpha
School Culture	10	0.888
Evidence-Based Practices	12	0.902
Ongoing Professional Development	6	0.866
Family Engagement	5	0.808

Table 2 shows the overall SIS reliability estimates compared across roles and grade levels with the number of participants in each group. Members of the implementation/leadership team rendered a coefficient alpha of 0.959 and non-members rendered an alpha of 0.969, showing that the reliability of the SIS was not severely impacted by involvement on the school leadership team who obtained additional training on the model and worked collaboratively to implement the model components within their school. Teachers produced a coefficient alpha of 0.960, administrators 0.929, other certified staff 0.953, and non-certified staff 0.970. The different grade levels revealed consistently high reliability with 14 elementary schools rendering a coefficient alpha of 0.955, four middle schools rendered 0.929, and five high schools rendered 0.968.

Table 2. Reliability Coefficient Estimates- Across Team Members, Staff Roles, and Grades

Implementation/Leadership Team	N	Cronbach's Alpha
Member	134	0.959
Non-Member	237	0.969
Staff Roles	**N**	**Cronbach's Alpha**
Teachers	294	0.960
Administrators	15	0.929
Other Certified Staff	24	0.953
Non-Certified Staff	11	0.970
Grade Levels	**N**	**Cronbach's Alpha**
Elementary	343	0.955
Middle School	128	0.929
High School	82	0.968

The overall SIS, essential elements domains, and implementation level sub-scales' reliability coefficients were compared to those of a different integrated RTI model in the first year of implementation at three urban charter schools with a total of 36 participants. The resulting Cronbach's alpha for the full SIS was 0.907, and element domains and implementation level sub-scales fell between 0.753 and 0.782 which meant that the SIS could produce reliable results across different integrated RTI models.

Table 3. Mean Comparisons for Implementation Team Members and Non-Members

Scale	Team Member (N=104)		Non-Team Member (N=237)		Significance
	Mean	Std. Dev.	Mean	Std. Dev.	
Full SIS	4.10	0.62	3.89	0.70	< 0.01
School Culture	4.21	0.62	3.95	0.74	0.08
Professional Development	4.13	0.73	3.90	0.80	0.02
Evidence-Based Practices	4.00	0.69	3.80	0.75	0.02
Family Engagement	4.10	0.63	3.96	0.75	0.20

The validity of the SIS was also examined by comparing the mean scores for various groups of participants and the means for implementation level sub-scales and essential element domains. Independent sample t-tests between implementation/leadership team members and non-team members showed that team members consistently rated implementation higher than did other school staff. Means for team members ranged from 4.00 to 4.21, but for non-members the means ranged from 3.80 to 3.96. These mean differences were significant for the overall SIS and the domains of Ongoing Professional Development and Evidence-Based Practices, as can be seen in Table 3. Eta squared effect size indicate that 2% of variance in SIS mean scores are accounted for by implementation/leadership team membership, and even though this is a small effect size results showed that “members of the school leadership team cannot accurately represent the perceptions of all staff” (Gaumer Erickson, et al., 2012, p. 44).

A One-Way Analysis of Variance (ANOVA) was conducted to compare means across the three levels of implementation (school, classroom, and individual student), and identified a significant difference between levels with $F(2,1035)=6.39$, $p<0.01$ and partial eta square of 0.012. Follow-up pair-wise comparison tests revealed that the Student Implementation Level ($M=3.84$, $SD=0.77$) was significantly lower than both the School Implementation Level ($M=4.03$, $SD=0.71$, $p=0.01$) and the Classroom Implementation Level ($M=3.99$, $SD=0.70$, $p=0.023$). To examine the mean differences across essential elements domains, another additional ANOVA procedure identified a significant difference with $F(3,1380)=3.46$, $p=0.016$ and a partial eta square of 0.007. Follow-up pair-wise comparison tests conducted between all pairs of implementation levels revealed that School Culture ($M=4.03$, $SD=0.71$) was significantly higher than Evidence-Based Practices ($M=3.86$, $SD=0.73$, $p=0.015$). This sequence of analyses indicated that the essential element domain of Evidence-Based Practices and the Individual Student Level of implementation were the measured components among the 14 schools that were least in place.

Implementation across years was examined by conducting a series of Independent t-tests between the 2010-2011 and 2011-2012 means for the full SIS, the four essential elements domains, and the three levels of implementation. The analysis included the results of 14 elementary, 4 middle, and 4 high schools that administered the scale during these two school years. Significant increases between 2010-2011 and 2011-2012 means were revealed for the full SIS, the essential elements domains of school culture, ongoing professional development, and family engagement; as well as the subscales of individual student level and school level implementation (see Table 4). These results showed that implementation of the integrated academic/behavior RTI model increased or improved from one year to the next. Delta (δ) effect sizes indicated that approximately 9-14% of the variance between the 2010-2011 and 2011-2012 results could be attributed to the passage of a full school years' worth of time.

Table 4. Mean Comparisons Across Years- Essential Elements and Implementation Levels

Scale	Year	N	Mean	Std. Dev.	Sig.	δ
Full SIS	2010-2011	784	3.81	0.72	0.02	0.12
	2011-2012	696	3.89	0.67		
School Culture	2010-2011	784	3.89	0.74	0.02	0.12
	2011-2012	696	3.97	0.69		
Professional Development	2010-2011	780	3.79	0.85	0.02	0.12
	2011-2012	694	3.89	0.76		
Evidence-Based Practices	2010-2011	781	3.71	0.81	0.08	0.09
	2011-2012	691	3.78	0.77		
Family Engagement	2010-2011	766	3.90	0.72	0.01	0.14
	2011-2012	674	4.00	0.68		

To further assess degree of implementation across years, participant ratings were converted to binomial scores where responses of 4 or 5 equaled implementation and responses of 1, 2, and 3 equaled lack of implementation. A threshold of 80% of school staff implementing each multi-tiered component was set as definition of school-wide implementation. This level is consistent with the research published by Sugai, Lewis-Palmer, and Todd in 1999, and has also been reported by Simonsen, Sugai & Negron (2008) as the sufficient level to indicate school-wide implementation of the model. For the 14 schools in their second year of implementation during 2010-2011, five items were found to be fully implemented and during 2011-2012 an additional seven items reached the threshold for full implementation among these schools.

The relationship between the School Implementation Scale results and academic achievement of students with disabilities were investigated by computing a Pearson Product Moment Correlation. Using the degree of implementation method describe above the percentage of school staff respondents that rated each item at 4 or 5 (threshold for school-wide implementation) were identified, and ranged from 1 item (3% of the total 33 items on the scale) to 30 items (91% of the items). The change between 2009 and 2011 in the percent of students with disabilities meeting proficiency on the Communication Arts state assessment for schools ranged from decreasing 17.5% to increasing 43.3%. The percent of the SIS items rated at level 4 or 5 by 80-100% of staff within each school was compared to the schools' increase in the percent of students with disabilities that met proficiency on the Communication Arts state assessment. A moderate correlation of 0.55 was computed between the two variables (Gaumer Erickson et al., 2012).

Reliability and Validity

The relationship among variables being researched and the psychometric criteria of reliability, convergent validity, discriminant validity, and construct validity, as well as among the psychometric criteria themselves are very complex (Peter & Churchill, 1986). Cook & Beckham (2006) claim that "a clear understanding of validity and reliability in psychometric assessment is critical for practitioners" (p.166.e14) because "validity and reliability relate to the interpretation of scores from psychometric instruments" (p.166.e7), and that "evidence should be sought to support a given interpretation. Reliable scores are necessary, but not sufficient, for valid interpretation" (p.166.e7). "The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations" (*Standards for Educational and Psychological Testing*, 1999, p.9), and "it is the interpretation of the scores required by proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each intended interpretation must be validated" (*Standards for Educational and Psychological Testing*, 1999, p.9).

The definition of reliability is the stability, consistency, and precision of responses; and instrument reliability includes measures of internal consistency such as item responses being consistent across constructs, test-retest correlations for stability of scores over time, and consistency in administration and scoring (Creswell, 2009). Reliability is the amount of random fluctuation in individual test scores, and is an important criterion of test quality because: 1) important decisions about individuals are based on test scores, but cannot be based on scores which are substantially random; 2) scores that are less than perfectly reliable cannot be perfectly valid, as scores can't both measure nothing and measure something; and 3) poor score reliability may compromise the ability of a study to yield noteworthy effects, due to the fact that perfectly unreliable scores are completely random and cannot yield significant results. According to Crocker & Algina (1986), reliability is concerned with score consistency and is relevant especially when there are ramifications for score interpretations because the more measurement error that exists in the scores, the less useful the scores may be for analysis and interpretation.

Henson (2001) described reliability as ratio of true score variance to observed score variance or the extent to which observed score variance is due to true score variance (degree to which scores are free of errors of measurement), and claim that reliability is critical when interpreting study effects and test results. Reliability Coefficients can be classified into the categories of stability, equivalence, precision, and internal consistency. The coefficient of stability is a correlation between scores from the same test form administered to same group on separate occasions. The coefficient of equivalence is the correlation between two different test forms administered to same examinees on the same occasion. The coefficient of precision is the theoretical correlation between test scores when examinees are administered the same test items repeatedly. The coefficient of internal consistency is used to estimate the coefficient of precision for a set of real test scores.

The most commonly used estimates of reliability are internal consistency coefficients (Henson, 2001) which can be readily calculated from a single administration of a test. Internal consistency coefficients measure performance across items using the inter-correlations among items, but measurement error can be caused by content sampling, guessing, or temporary fluctuations of individual performance. Item homogeneity exists when examinees perform consistently, which provides evidence that the items measure the same construct in approximately the same way. Internal consistency coefficients are not direct measures of reliability but are theoretical estimates derived from Classical Test Theory because they are related to item homogeneity or the degree to which the items on a test jointly measure the same construct. This allows researchers to interpret the composite score as a reflection of all the test's items, and is important in both substantive and measurement contexts (Henson, 2001). There are several internal consistency reliability estimates and they correspond to the three sources of measurement error: content sampling of items, stability across time, and inter-rater reliability. Content sampling is the theoretical idea that a test is made up of a random selection of all possible items (which are highly interrelated because they are designed to assess the same construct), and the interrelationship between items is typically called item consistency.

"Scale reliability is the proportion of variance attributable to the true score of the latent variables" (DeVillis, 2003, p. 27), and internal consistency reliability is concerned with the homogeneity of the items within the scale (DeVillis, 2003). This internal consistency is typically equated with the widely used measure of reliability- Cronbach's (1951) coefficient alpha. Coefficient Alpha (Cronbach, 1951) allows one to estimate the reliability of a composite based on the composite variances and covariances because the function of the variance of the composite scores and covariances of the tests that make up the composite. This portion of uniqueness variance and covariance can be calculated using the Kuder-Richardson-20 formula (Richardson & Kuder, 1939), but Cronbach's coefficient alpha is a more general form of the KR-20 formula which can be used with dichotomously scored items or with measures using multiple response categories such as Likert scale data (Henson, 2001). Coefficient alpha invokes the general linear model to calculate the ratio of explained to total variance which must also account for the intercorrelations among items under the assumption that as items are more highly correlated, the magnitude of alpha will increase.

The computation of coefficient alpha partitions the total variance among the set of items into actual variation and error, and alpha is defined as "the proportion of a scale's total variance that is attributable to a common source, presumably the true source of a latent variable underlying the items" (DeVillis, 2003, p. 31). The formula for Cronbach's coefficient alpha is:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\Sigma\sigma^2{}_i}{\sigma^2{}_{yi}}\right)$$

This equation can be summarized by stating that a "measure's reliability equals the proportion of total variance among its items that is due to the latent variable and thus is communal" (DeVillis, 2003, p. 35), and "the formula for alpha expresses this by specifying the portion of total variance for the item set that is unique, subtracting this from one to determine the proportion that is communal, and multiplying by a correction factor to adjust for the number of elements contributing to the earlier computations" (DeVillis, 2003, pp. 35-36). For tests that consist of different scales measuring different constructs, internal consistency should be measured separately for each scale. Some researchers have stated that reliabilities of 0.60 or 0.50 will suffice, but Nunally (1978) suggested that 0.80 was adequate for basic research purposes and 0.90 would be minimally tolerable for applied settings with 0.95 being the desired level.

There are salient data features that affect coefficient alpha such as perfectly uncorrelated items (because the items share no variance such that covariances and correlations are 0 and there is no internal consistency among the items), perfectly correlated items (when the items possess perfect internal consistency and alpha reaches maximum of 1), and perfectly correlated items with mixed signs (items are highly correlated but not all in the same direction which causes a "paradox" in the calculation of alpha). If one knows the proportion of true score variance, one then also knows the proportion of error variance; and knowing the error variance allows researchers to state their confidence that an examinee's test score accurately reflects their abilities or perceptions of the construct being measured.

DeVillis (2003) explained that a scale's alpha is influenced by two characteristics: the extent of the covariation among the items and the number of items on the scale itself. He noted the effect of the length of a scale on its reliability, and explained that while shorter scales place less of a burden on respondents, longer scales are good because they tend to be more reliable. Therefore, in the construction of a scale there is a trade-off between brevity and reliability. Whether dropping "bad" items actually increases or slightly lowers the coefficient alpha depends on how poor the items are that will be dropped, and on the number of items remaining on the scale. Items that contribute least to the overall internal consistency of the scale should be the first to be considered for exclusion from the measurement instrument. Holding the number of scale items constant, reliability will increase as the sum of the item variances decrease and the total score variance increases. Classical reliability estimates rely on the variance of the total scores.

Different samples, testing conditions, or changes in any other factor may affect observed scores and therefore affect the reliability estimates- yielding different reliabilities because the total variance is likely to change (Henson, 2001). The sources of variance in classical test theory are separate and cumulative, and the effect size magnitude is inherently attenuated by the reliability of the scores used to obtain the effect estimate and can also affect statistical power (Henson, 2001). Thus, Henson (2001) suggested that reliability should always be reported and considered in result interpretation, and that researchers should report reliabilities for their scores and not depend on prior studies or test manuals. According to Gronlund & Linn (1990), reliability refers to the results obtained with an evaluation instrument and not to the instrument itself, and they believe it is more appropriate to speak of the reliability of test scores or the measurement than of the test or the instrument. In 1999, the APA Task Force on Statistical Inference recommended authors provide reliability coefficients and mandated that authors always report effect sizes for primary outcomes because interpreting the size of observed effects requires as assessment of the reliability of the scores (Wilkinson & APA Task Force on Statistical Inference, 1999). DeVillis (2003) stressed that procedures and samples used in reliability studies should be sufficiently described to permit users to determine similarity between conditions of the reliability study and their local situations. When a test is normally used for a particular population of examinees reliability estimates and standard errors of measurements should be reported separately for such populations, and when test scores are used primarily for describing or comparing group performance reliability and standard errors of measurement for aggregated observations should be reported. Reliability also affects statistical power- all else being constant, poor reliability will reduce the power of statistical significance tests (when effects are reduced they are more difficult to detect) so poor measurement leads to potential problems with statistical analysis (Henson, 2001). Reliability of the scores is central to understanding the observed relationships between the variables being studied.

Shadish, Cook, and Campbell (2001) defined validity as well-grounded, just, and producing the desired results. "Validity concerns the degree to which scores reflect the intended underlying construct, and refers to the interpretation of results rather than the instrument itself," and is "best viewed as a carefully structure argument in which evidence is assembled to support and refute proposed interpretations of results" (Cook & Beckham, 2006, p.166.e14).

The perspectives and beliefs regarding validity theory in psychological and educational measurement have changed over time. In 1989, Messick identified and began to stress other validity concerns such as interpretation and use of measurement results. Interpretation refers to describing the examinee's level on the construct of interest and use refers to the decisions that are made on the basis of those scores. A new unified concept of validity was introduced by Messick which inter-related these issues as fundamental aspects of a more comprehensive theory of construct validity that addressed both meaning and social values in test interpretation and test use. Messick redefined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on the test scores or other modes of assessment" (1990, p. 1). He agreed that validity was the extent to which a test actually measures what it purports to measure, as well as the extent to which score meanings and action implications hold across persons or groups and across settings or contexts. Every measurement issue is a matter of validity, and any limits in the interpretability and use of test scores are related to problems associated with their validity.

Messick (1995) claimed that validity was a unitary concept with various types of validity evidence, and that the types of evidence necessary for each measurement instrument depend upon the purpose of assessment and the types of evidence necessary to support that purpose. He described six types of construct validity, including consequential validity which concerns the impact on respondents of how their scores are used. The six distinguishable aspects of construct validity (content, substantive, structural, generalizability, external, and consequential) are the means for addressing central issues implicit in the notion of validity as a unified concept, and the aspects function as general validity criteria or standards for all educational and psychological measurement (Messick, 1995).

The content aspect of construct validity includes evidence of content relevance, representativeness, and technical quality. The substantive aspect refers to theoretical rationales for the observed consistencies in test responses along with empirical evidence that the theoretical processes are actually engaged in the assessment task. The structural aspect appraises the fidelity of the scoring structure of the construct domain at issue. The generalizability aspect examines the extent to which score properties and interpretations generalize to and across population groups, settings, and tasks. The external aspect includes convergent and discriminant evidence from multi-trait multi-method comparisons as well as evidence of criterion relevance and applied utility. The consequential aspect appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice. These evidence-based sources of construct validity provide the theoretical rationale underlying score interpretation, construct validation on internal and external test structures, correlations between test scores and criterion measures, and social consequence of test interpretation and use. The construct validity of score interpretation provides the basis for all score-based inferences, not just those related to interpretive meaningfulness but also the content and criterion related inferences specific to applied uses (Messick, 1995).

The current conceptualization of validity no longer refers to different types of validity but instead to different lines of validity evidence, and all validity evidence serves to provide information relevant to a specific intended interpretation of test scores. The essence of unified validity is that the appropriateness, meaningfulness, and usefulness of score-based inferences are inseparable- providing meaning and value to test validation. The integrating power derives from empirically grounded score interpretation which is supportive of score meaning or consequences, and contributes to score validation and function of the outcome of the testing for either interpretation or applied use.

Validity is an integrative summary of evidence pertaining to all six aspects, and together they provide a way of addressing the multiple and interrelated validity questions that need to be answered to justify score interpretation and use (Messick, 1995). Value judgments are likely to be associated with test interpretation and must be explored and communicated. Relevance of scores in applied settings is necessary for making intended decisions, and utility of scores relates to the benefits of testing relative to its cost. Social consequences of the test scores, the intended outcomes of test use and their potential side effects, also require examination as validity evidence (Messick, 1995).

Validity, however, is an approximate truth of an inference that requires evidence to support the inference as being true or correct (Shadish, Cook, & Campbell, 2001). Validity is a property of inferences, and there are no methods that can guarantee the validity of an inference. "The essential question of test validity is how well a test does the job it is employed to do. The same test may be used for several different purposes, and its validity may be high for one, moderate for another, and low for a third" wrote Cureton in 1951 (p. 621). Cronbach and Thorndike reiterated this sentiment in 1971 by emphasizing that validity was not inherent within a test, but must be evaluated for each different application of a test.

Cronbach and Thorndike (1971) described validation as the process by which a test developer or test user accumulates evidence (theoretical basis, construction, scoring, interpretation, use, reliability, and predictive power, for example) to support the types of inferences that are to be drawn from the test scores. The validity of a scale, however, is not firmly established during scale development; validation is a cumulative ongoing process. "Validity is a characteristic of how a scale is used, not of the scale itself" (DeVellis, 2003, p.159). Without the knowledge of the extent to which the validity of the scale generalizes across populations, settings, specific details of administration, or an assortment of other dimensions, its use is limited. Identical instrument performance cannot be assumed across settings and populations, and researchers must address the possibility of the presence of differential functioning and the limitations it may impose on their conclusions. "Any conclusions based on a scale that has had limited use should consider the following: (a) how its present application differs from the context of its original validation, (b) the likelihood that those differences might limit the validity of the scale, and (c) the implications of those limitations for the present research" (DeVellis, 2003, p.159).

Sources of validity evidence for use of an instrument in a specific context can be based on test content, on test takers' response processes, on the instrument's internal structure, on the consequences of the testing or decisions made from test results, or on the relations of the test and its scores to others variables. These relationships to other variables can be convergent (measures of related constructs), divergent (measures of different constructs), test-criterion (other measure of the same construct), or validity generalization. All of these sources of validity evidence are then integrated into a coherent argument in favor of the use of the measurement instrument for the intended use and score interpretation of the context in which it is to be employed. "A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of the test scores for specific uses. It encompasses evidence gathered from new studies and evidence available from earlier reported research" (The Standards for Educational and Psychological Testing, 1999, p.17).

Measurement Invariance

According to Vandenberg & Lance (2000), "measurement can be defined as the systematic assignment of numbers on variables to represent characteristics of persons, objects, or events" (p.4), and in education, measurement is "typically aimed at describing characteristics of individuals, groups, or organizations that are of some substantive interest" (p.4) to those administering the measurement. However, Widaman & Reise (1997) claimed that the definition of measurement invariance- that the relations of latent variables with their factor indicators must be identical across groups- is a broad one. "A measure is invariant when members of different populations who have the same standing on the construct being measured receive the same observed score on the test" (Schmitt & Kulijanin, 2008, p.211), and invariance testing is the process of investigating the invariance across groups of relations among psychological measures (Widaman & Reise, 1997, p.281). It is important that the instrument demonstrate construct invariance over the factors being measured in order to recommend the instrument's use for other populations or in other contexts.

Much discussion has been devoted to this topic, and many researchers are often cited as providing the theoretical rationale for tests of measurement equivalence/invariance (ME/I) (Byrne, Shavelson, & Muthen, 1989; Cole & Maxwell, 1985; Dragsow & Kanfer, 1985; Horn & McArdle, 1992; Joreskog, 1974; Millsap & Everson, 1991; Millsap & Hartog, 1988; Widaman & Reise, 1997; Rock, Werts, & Falugher, 1978). Invariance testing involves investigating the invariance across groups of relations among psychological measures (Widaman & Reise, 1997), and "the general question of invariance of measurement is one of whether or not, under different conditions of observing and studying phenomena, measurements yield measures of the same attributes" (Horn & McArdle, 1992, p.117). For test scores to be comparable across distinct examinee populations, the observed test items or indicators, must have identical or invariant, quantitative relationships with the latent variable for each population of interest (Widaman & Reise, 1997; Meredith, 1993). When tests or assessment instruments meet this criterion, they are said to display measurement invariance.

Little (1997) described two types of factorial or measurement invariance. The first type of invariance concerns the psychometric properties of the measurement scales which includes configural invariance (Buss & Royce, 1975; Suzuki & Rancer, 1994), metric invariance (Horn & McArdle, 1992; Meredith, 1993), measurement error invariance (Mullen, 1995; Singh, 1995), and scalar invariance (Meredith, 1993; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). The second type of invariance concerns between-group differences in latent means, variances and covariances. Category 1 invariance is a prerequisite for the interpretation of Category 2 differences, whereas Category 2 differences are usually the data having substantive research interest (Little, 1997). Invariance of uniqueness or residual variances is referred to by Meredith (1993) as “strict invariance” and is recognized by most researchers as difficult to achieve. However, many researchers do not believe strict invariance is necessary in order to test differences in factor structure or latent means, which are the substantive questions most often of interest to researchers (Widaman & Reise, 1997; Stewart, 2006).

If the goal of the researcher is to compare group means on the basis of observed test scores, bias can be a serious problem (Borsboom, 2006). Unless biasing effects in the groups being compared cancel each other out, then mean group differences in observed scores may not reflect differences in the latent variables of interest because the observed scores are confounded, and the size of the biasing effects is crucial (Borsboom, 2006). If one’s research interest is not to compare means between groups, but rather to investigate how variables are related within different groups, then bias may be entirely irrelevant. Whether the biasing effect is actually a validity threat partly depends on the source of the biasing effects (Borsboom, 2006). If the source of the biasing effect is that item responses do depend on a second latent attribute not targeted by the researcher that the groups possess in uneven amounts, the situation is different and then the test scores depend on more than the latent variable of interest and test scores are likely to violate unidimensionality in each of the groups as well as in a multigroup analysis (Borsboom, 2006). According to Borsboom (2006), when tests are used in the selection of individuals, rather than for scientific research on population characteristics, they should conform to higher psychometric standards because of the danger of bias. Therefore, when the purpose of test use is for the selection of individuals, measurement invariance is a necessary condition for fair selection procedures (Borsboom, 2006). "Whether bias is to be considered an important validity threat is not a straightforward function of p-values or effect sizes. The reason for this is that the importance of bias is partly a pragmatic issue: It depends on aspects of the research situation that are not statistical in nature, such as the purposes for which the test scores are being used" (Borsboom, 2006, p.177). Further, if the biasing factors have effects both within and between groups, then relations found within groups may be confounded as mean differences between groups so that bias again becomes a validity threat (Borsboom, 2006).

Windle, Iwawaki, and Lerner explained in 1988 that the primary method for assessing measurement invariance has been through the use of factor analytic techniques, and "involves the study of similarities and differences in the covariation patterns of item-factor relations" (Widaman & Reise, 1997, p.551). Widaman and Reise claimed in 1997 that factor analytic models most commonly invoked when discussing issues related to measurement invariance, but this has changed with the advent of more sophisticated methods for confirmatory modeling of item-latent variable relations using Confirmatory Factor Analysis (CFA) (Joreskog, 1971) and item response theory (IRT) models (Lord, 1980). Vandenberg and Lance stated in 2000 that

> "historically, evaluation of measurement quality has been rooted in Classical Test Theory (CTT) of true and error scores (Crocker & Algina, 1986; Lord, Novick, & Birnbaum, 1968; Nunnally & Bernstein, 1994). CTT has provided and probably will continue to provide a solid foundation for the evaluation of manifest or observed variables; measurement properties in terms of reliability and validity. However, additional issues extend beyond the traditional purview of CTT that represent important considerations in evaluating manifest variable's measurement properties, and relatively recent advances in analytic tools

have made investigation of these issues much more accessible to researchers" (p.5).

There are questions underlying measurement equivalence/invariance (ME/I) that are not directly accessible through CTT avenues and after a fairly extensive review of measurement equivalence/invariance literature, Vandenberg and Lance (2000) claimed that a CFA framework was the most common and effective method for testing measurement invariance for multi-item composite measures.

Currently there is a general agreement among researchers as to the specific methods of conducting various tests of ME/I- through constraining factor loadings to be equal across groups to test metric invariance- but there has been little consensus among sources as to the set of tests that constitute a thorough evaluation of ME/I or on the sequence of tests that should be conducted (Vandenberg & Lance, 2000). There has been general agreement among researchers that an omnibus test of equality of covariance matrices across groups was an important first step, and that a test of configural invariance is necessary and can serve as a baseline model for further tests (Vandenberg & Lance, 2000). Vandenberg and Lance (2000) summarized recommendations made by authors of articles where the main intent was to provide a theoretical rationale and describe the procedures for testing ME/I in a CFA framework, and this Summary of Recommended Practices outlines a practical ordering for the necessary measurement equivalence/invariance tests using multi-sample applications (p.12).

First is an omnibus test of the equality of covariance matrices across groups to test the null hypothesis of invariant covariance matrices of different groups. Second is a test of configural invariance or test of the weak factorial invariance null hypothesis (Horn & McArdle, 1992) in which the same pattern of fixed and free factor loadings is specified for each group.

Configural invariance must be established in order for subsequent tests to be meaningful. Third is a test of metric invariance (Horn & McArdle, 1992) or test of the strong factorial invariance null hypothesis where factor loadings for like items are invariant across groups, and at least partial metric invariance must be established in order for subsequent tests to be meaningful. Fourth is a test of scalar invariance (Meredith, 1993; Steenkamp & Baumgartner, 1998) or test of the null hypothesis that intercepts of like items' regressions on the latent variables are invariant across groups. Fifth is a test of the null hypothesis that like items' unique variances are invariant across group. Tests of scalar invariance should only be conducted if at least partial metric invariance is established, and tests of invariant uniqueness should proceed only if at least partial metric and scalar invariance has been established. Sixth is a test of the null hypothesis that factor variances are invariant across groups, seventh is a test of the null hypothesis that factor covariances are invariant across groups, and eighth is a test of the null hypothesis of invariant factor means across groups as a way to test for differences between groups in level on the construct of interest. Vandenberg and Lance (2000) maintained Byrne et al.'s (1989) distinction that the first 5 tests are tests of aspects of measurement invariance (as they concern tests of relationships between measured variables and latent constructs), and the next 3 tests were testing aspects of structural invariance (as they refer to tests concerning the latent variables themselves).

"Measurement invariance is usually tested using Multi-group Confirmatory Factor Analysis, which examines the change in the goodness-of-fit index (GFI) when cross-group constraints are imposed on a measurement model" (Cheung & Rensvold, 2002, p.233). Multi-group Confirmatory Factor Analysis (MGCFA) is an extension of CFA which tests the invariance of estimated parameters of two nested models across groups, and the degree of invariance is most frequently assessed by the Likelihood Ratio Test or the difference in χ^2

between two nested models. However, according to Cheung and Rensvold in 2002, there were no generally accepted criteria in MGCFA for determining if changes in the practical GFIs were meaningful when measurement constraints are added. There was also a lack of consensus about what constitutes significant GFI differences, and this placed limits on measurement invariance testing.

"A model is considered suitable if the covariance structure implied by the model is similar to the covariance structure of the sample data, as indicated by an acceptable value of the goodness-of-fit index (GFI)" (Cheung & Rensvold, 2002, p.234). The most commonly used GFI is the χ^2 statistic. Nonsignificant values of χ^2 indicate failure to reject the null hypothesis that the hypothesized covariance matrix is identical to the observed covariance matrix which is usually accepted as evidence of adequate fit (Cheung & Rensvold, 2002). Many different GFIs have been proposed as alternatives to χ^2. These other GFIs include the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), Normed Fit Index (NNFI; Bentler & Bonett, 1980), and root mean squared error of approximation (RMSEA; Steiger, 1990). However, since these GFIs have no known sampling distributions, researchers have proposed many various criterion values indicative of satisfactory model fit.

Cheung & Rensvold (2002) assessed 20 GFIs based on the minimum fit function, and their simulation study examined changes in the GFIs (ΔGFIs) under the two group situation when invariance constraints were added. The effects of sampling error and model characteristics on MGCFA outcomes (differences in GFIs/ΔGFI obtained when an unconstrained model is compared with one having measurement invariance constraints under the null hypothesis of invariance) were tested. Critical values of ΔGFIs independent from model characteristics and based on the sampling distributions of ΔGFIs are obtained using simulations (Cheung &

Rensvold, 2002). Based on their results, the Δ comparative fit index, Δ Gamma hat, and Δ McDonald's Noncentrality Index were recommended to evaluate measurement invariance (p.233). Cheung and Rensvold (2002) concluded that the "results show that ΔCFI, Δ Gamma hat, and Δ McDonald's NCI are robust statistics for testing the between-group invariance of CFA models" (p. 250), and "although the standard errors and critical values differ for the different invariance models, the between-model variations are so small that a general criterion for all hypotheses can be proposed" (p. 251). They reported that a value of ΔCFI smaller than or equal to –0.01 indicated that the null hypothesis of invariance should not be rejected. For ΔGamma hat and ΔMcDonald's NCI, the critical values were –.001 and –.02, respectively (Cheung & Rensvold, 2002).

Chen (2007) examined the sensitivity of goodness-of-fit indexes to lack of measurement invariance with 2 Monte Carlo studies for 3 commonly tested levels: factor loadings, intercepts, and residual variances. He concluded that the standardized root mean squared residual (SRMR) appeared to be more sensitive to lack of invariance in factor loadings than in intercepts or residual variances (Chen, 2007). Chen also found that CFI and SRMR appeared to be equally sensitive to all three types of lack of invariance, and that changes in statistics are affected by the interaction between the pattern of invariance and the proportion of invariant items (2007). Unequal sample sizes affected changes across all three levels of invariance and changes were bigger when sample sizes were equal rather than when they were unequal (Chen, 2007).

"The establishment of measurement invariance across groups is a logical prerequisite to conducting substantive cross-group comparisons, but measurement invariance is rarely tested in organization research" (Vandenberg & Lance, 2000, p.4). According to Borsboom (2006), "it appears that measurement invariance is rarely explicitly investigated. Many researchers simply assume their measures to be invariant across groups without checking this assumption" (p.180), and continued to explain that this could be problematic because even subtle violations of measurement invariance may lead to spurious conclusions. Dragsow (1984, 1987) stated that to compare groups of individuals on their level on a trait or to investigate whether trait-level scores had differential correlates across groups, the researcher must assume that the numerical values under consideration are on the same measurement scale. Thus, the researcher must assume that "measurement invariance" holds across groups. If trait scores are not comparable across groups, then differences between groups in mean levels or in the pattern of correlations of the test with external variables could be potentially artificial and substantively misleading (Meredith, 1993). "Differences in raw scores of different groups cannot be used to infer group differences in theoretical attributes unless the test scores accord with a particular set of model invariance restrictions. Statistically, this means that the mathematical function that relates the latent variables to the observations must be the same in each of the groups involved in the comparison" (Borsboom, 2006, p.176). Due to presently available methodological techniques, Borsboom stated that this situation need not continue and "investigating measurement invariance should now become a routine part of research into the structure of group differences" (2006, p.180).

Professional Standards

"Using tests or instruments that are valid and reliable to measure [abstract] constructs is a crucial component of research quality" (Kimberlin & Winterstein, 2008, p. 2283), and extremely critical when policy decisions may follow from the outcomes and recommendations of the evaluation results for an educational program or intervention. *The Program Evaluation Standards* (2011) and the *American Evaluation Association Guiding Principles* (2004) are sets of professional standards that promote ethical practices in evaluation to ensure quality methods and valid and reliable outcomes. Educational evaluators must follow these standards to ensure that an evaluation will a) serve the information needs of intended users; b) be realistic, prudent, diplomatic, and frugal; c) be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results; and d) reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. In order to ensure that the specified *Program Evaluation Standards* (2011) and *AEA Guiding Principles for Evaluators* (2004) can be met regarding evaluation research using the School Implementation Scale, *The Standards for Psychological and Educational Testing* (2014) will be used as a guide for an in-depth psychometric analysis of the School Implementation Scale and its use with different populations and across various integrated academic and behavior RTI models.

The Standards for Educational and Psychological Testing were created to "provide a basis for evaluating the quality of testing processes" (*Standards for Educational and Psychological Testing*, 1999, p.1) and the effects or consequences of test use; and to "provide a frame of reference to assure that relevant issues are addressed" (p. 2). The standards are separated into three sections entitled "Foundations," "Operations," and "Testing Applications." The Foundations section focuses on fundamental testing issues such as validity, reliability, and fairness. The Operations section deals with operational testing issues such as test design and development, administration, scoring reporting, and supporting documentation recommended for tests. The Testing Applications section details specific applications in testing, such as workplace testing and credentialing, educational testing and assessment, and the use of tests for program evaluation, policy studies, and accountability (Ernesto, 2013). The primary purpose of the testing standards is providing criteria for evaluating tests and testing practices (Ernesto, 2013).

"Depending on the context and purpose of test development or use, some standards will be more salient than others. Moreover, some standards are broad in scope, setting forth concerns or requirements relevant to nearly all tests or testing contexts, and other standards are narrower in scope" (The Standards for Educational and Psychological Testing, 1999, p.2). It is recommended that each standard be considered and that its applicability to the testing context under consideration be determined, but there is no expectation that evidence be collected related to each standard. "Evaluating acceptability involves a professional judgment that is based on a knowledge of behavioral science, psychometrics, and the community standards in the professional field to which the tests apply; the degree to which the standard has been satisfied by the test developer and user; the alternatives that are readily available; and research and experiential evidence regarding feasibility of meeting the standard" (The Standards for Educational and Psychological Testing, 1999, p.4).

However, when the measurement instrument or scale is being used for evaluation purposes, The Program Evaluation Standards (2011) and the AEA Guiding Principles for Evaluators (2004) become crucial. One main goal of evaluation research, such as studies using the School Implementation Scale to evaluate integrated academic and behavior RTI models, is that it responds to the stakeholders' questions and concerns regarding that which is being evaluated, and the evaluation results can then be used to inform decisions regarding that program or intervention. Professional standards are a critical part of program evaluation because of the potential impacts to programs and individuals. Because the SIS is designed to be used in program evaluation efforts, those who administer or use it should be held to the standards intended for guiding program evaluators. The Program Evaluation Standards (2011) were created by the Joint Committee on Standards for Educational Evaluation. The goal of these standards is to help users recognize and improve evaluation quality (p. xxvii), and each of five parts present one attribute of quality and its accompanying standards.

Part I of The Program Evaluation Standards (2011) is Utility, which includes eight standards and discusses use, usefulness, influence, and misuse of evaluation information. It describes when and how evaluation worth is created, for example, when evaluations contribute to stakeholders' learning, inform decisions, improve understanding, lead to improvements, or provide information for accountability judgments (p. xxviii). Part II is Feasibility, which includes four standards and discusses the effects of contexts, cultures, costs, policies, power, available resources, and other factors on evaluations. It details which feasibility factors to consider before implementing an evaluation and how to increase or maintain feasibility in different contexts (p. xxviii). Evaluations are feasible when they can take place with an adequate degree of effectiveness and efficiency. Part III of the standards is Propriety, which includes seven standards and considers the rights of stakeholders and other persons and details the responsibilities of all stakeholders, especially evaluation professionals, in an evaluation. Setting thresholds for adequate propriety can be difficult and requires balancing different stakeholders' needs and situations. Social justice considerations may play a significant role, but not all stakeholders have the same values or conceptions of social justice (p. xxviii). Part IV of The Program Evaluation Standards (2011) is Accuracy, which includes eight standards regarding how to increase the accuracy of findings and conclusions; and discusses reliability, validity, reduction of error and bias. Accuracy addresses quality in data collection, analysis, logic, conclusions, and communication (p. xxviii). Accuracy is the truthfulness of evaluation representations, propositions, and findings- especially those that support judgments about the quality of program or program components. Part V is Evaluation Accountability, which includes three standards pertaining to internal and external metaevaluation providing the methodology used to increase and document evaluation quality (p. xxviii).

According to The Program Evaluation Standards (2011), judgments about an evaluation's utility are made based on the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs. Understanding evaluation utility is to examine the variety of possible users for evaluation processes, findings, and products (Program Evaluation Standards, 2011, p. 4). The goal of the utility standards is to increase the likelihood that the evaluation will have positive consequences and substantial influence (p. 8). The program evaluation standards of utility are related to the AEA Guiding Principles for Evaluators (2004) of integrity and honesty and of responsibilities for the general and public welfare. The integrity and honesty principle states that the evaluator should not misrepresent their procedures, data, or findings. The responsibilities for general and public welfare principle states that evaluators should take into account the diversity of interest and values of the general public, and should consider not only the immediate outcomes of that which is being evaluated but also its broad assumptions, implications, and potential side effects. If the evaluator, whether intentionally or unintentionally, misrepresents the procedures or results of the study, its utility is limited. Additionally, if the interests and values of the stakeholders and general public, or if the long-term outcomes of the study such as implications or potential side effects are not considered, the utility of the evaluation research is compromised.

Another necessity of evaluation research is that it be feasible within the context of the program or intervention being evaluated. Four key concepts are relevant to the discussion of evaluation feasibility: (a) Evaluability, or the degree to which it is possible to evaluate a specific program at a specific time and place; (b) Context, or the cultural, political, economic, governmental, and geographical circumstances and environments in which the program occurs; (c) Values, or the systems of concepts and qualities that stakeholders use to prioritize and judge aspects of the lives; and (d) Accountability, or resource use; effective and efficient use of evaluation and program resources, including time, money, and people, to create value (Program Evaluation Standards, 2011, p. 72-74). The context, values, and accountability aspects of the feasibility standards for program evaluation are especially pertinent for the current study. These feasibility program evaluation standards are akin to the AEA Guiding Principles for Evaluators of respect for people and respect for general and public welfare. The respect for people principle states that evaluators should seek a comprehensive understanding of the important contextual elements of the evaluation that may influence the results. The lack of evaluator attention to the context of that which is being evaluated, the values of the individuals judging that entity, or the accountability of using resources effectively and efficiently can all decrease the feasibility of the evaluation research. Feasibility is also reduced by focusing on the immediate outcome, and not considering the broad assumptions, implications, and potential side effects of the evaluation research as written in the respect for general and public welfare principle.

The most important quality of evaluation research is that it needs to be accurate, especially in light of judgments and decisions that could possibly be made about that which is being evaluated. Accuracy is achieved through sound theory, methods, designs, and reasoning. Evaluations should strive for as much accuracy as is feasible, proper, and useful to support sound conclusions and decisions in specific situations. One goal of the accuracy standards is to point out the specific components of an evaluation that should be accurate. The eight accuracy standards focus on what it means for specific components to be accurate and how to increase their accuracy (Program Evaluation Standards, 2011, p. 158). Accuracy of the following evaluation components is especially important:

- Findings, interpretations, conclusions, extrapolations, and decisions (A1 & A2)
- Reasoning, including the supporting theoretical frameworks (A5, A6, & A7)
- Concepts and terms (A2 & A8)
- Information and analyses (A2, A3, A5, & A6)
- Descriptions of programs, program theoretical frameworks, and their contexts (A4)
- Communication and reporting (A8)

Taken together, the accuracy standards also point out how to minimize factors that can undermine accuracy in evaluation through inconsistencies, distortions, and/or misconceptions (Program Evaluation Standards, 2011, p. 158).

Another set of professional standards intended for program evaluators are the American Evaluation Association (AEA) Guiding Principles for Evaluators (2004). AEA strives to promote ethical practice in the evaluation of programs, products, personnel, and policy; and developed these Principles to guide evaluators in their professional practice. The Guiding Principles are an abbreviated version of the full principles that were developed and endorsed by the American Evaluation Association in 1994 and reviewed and revised (and ratified) in 2004. The program evaluation standards of accuracy are very much related to the AEA Guiding Principles for Evaluators of systematic inquiry and integrity and honesty. The principle of systematic inquiry states that evaluators should conduct systematic, data-based inquiries about that which is being evaluated while adhering to the highest technical standards of the methods used and make clear any limitation of an evaluation and its results. Contextual values, assumptions, theories, methods, results, and analyses significantly affect the interpretation of evaluation findings; and if the evaluation research lacks sufficient attention to these aspects, the accuracy of the results can be highly compromised. Accuracy of the evaluation research can also be impacted by evaluators misrepresenting their procedures, data, or findings as written in the integrity and honesty principle.

These principles are broadly intended to cover all kinds of evaluation; however, the common ground is that evaluators aspire to construct and provide the best possible information that might bear on the value of whatever is being evaluated. The five principles proposed in the AEA document are not independent, but can overlap in many ways and sometimes they may conflict. For that reason, evaluators need to review them carefully, and choose the most appropriate or vital among them for a particular situation. These principles are not intended to replace standards by evaluators or by the other disciplines in which evaluators participate. These principles are part of an evolving process of self-examination by the profession, and should be revisited on a regular basis. The five AEA Guiding Principles for Evaluators (2004) consist of:

1) Systematic Inquiry, in which evaluators should conduct systematic, data-based inquiries about whatever is being evaluated while adhering to the highest technical standards appropriate to the methods they use and making clear the limitations of an evaluation and its results. Contextual values, assumptions, theories, methods, results, and analyses significantly affecting the interpretation of the evaluative findings should be discussed.
2) Competence, which states that evaluators provide competent performance by possessing the education, abilities, skills, and experience appropriate to undertake the tasks proposed in the evaluation. Evaluators should also be culturally competent to ensure recognition, accurate interpretation and respect for diversity.

3) Integrity and Honesty, which ensure the honesty and integrity the evaluator's own behavior and of the entire evaluation process including negotiating, disclosing any potential conflicts of interest, and not misrepresenting their procedures, data, or findings.
4) Respect for People, which guide evaluators to respect the security, dignity and self-worth of respondents, program participants, clients, and other stakeholders with whom they interact and maximize the benefits while reducing any unnecessary harms that might occur. Evaluators should also seek a comprehensive understanding of the important contextual elements of the evaluation that may influence the results.
5) Responsibilities for General and Public Welfare, in which evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare, and should consider not only the immediate operations and outcomes of whatever is being evaluated, but also its broad assumptions, implications, and potential side effects.

The utility, feasibility, and accuracy standards from the 2011 Program Evaluation Standards and the systematic inquiry, integrity and honesty, respect for people, and respect for the general and public welfare standards from the 2004 AEA Guiding Principles for Evaluators seem to be most salient for framing the current study examining the psychometric properties of the School Implementation Scale instrument currently being used in program evaluation efforts. The Standards for Educational and Psychological Testing (2014) will be the guide for gathering additional validity and reliability evidence related to the SIS's continuing use with different populations and across various integrated academic and behavior RTI models, specifically for evaluation purposes. In order to ensure that the specified Program Evaluation Standards (2011) and AEA Guiding Principles for Evaluators (2004) can be met regarding evaluation research using the School Implementation Scale, The Standards for Psychological and Educational

Testing (2014) will be used as a guide for an in-depth psychometric analysis of the SIS and its results. There are important psychometric issues related to the item-factor structure and construct invariance of the SIS, the relationship between the SIS factors/latent domains, and the impact of various characteristics from the individual respondent level to the school, district, and state level or type of RTI model that were not addressed in the 2012 Gaumer Erickson et al. SIS study.

CHAPTER 3

Methods and Procedures

Following the most pertinent and critical of The Program Evaluation Standards (2011) and the American Evaluation Association Guiding Principles for Evaluators (2004) for measuring implementation of multi-tiered systems of support, and using the *Standards of Education and Psychological Testing* (2014) as a guide, this study examined the psychometric properties of the School Implementation Scale for use in evaluative studies of different integrated academic and behavior tiered RTI models. There were important psychometric issues requiring more sophisticated data analysis procedures that had not yet been investigated. These psychometric were related to the theory-driven four-factor essential elements of successful school structure, and the measurement invariance of the School Implementation Scale and its constructs across states with different overall tiered response-to-intervention models and across the major types of integrated academic and behavior tiered response-to-intervention models implemented in those states.

The general version of the School Implementation Scale currently contains 31 Likert-type items, and four respondent demographic items: school name, staff role, number of years at the school, and membership on the school leadership team. However, the SIS is customizable to meet the needs of each state by eliminating some items and adding some of their own items. Of the 31 SIS items, the two states in which data was collected included 28 of the items in each dataset. These 28 common items, which are the non-shaded items in Appendix A, were used for analyses. Staff role response options include administrator, general education teacher, special education teacher, other certified staff, and non-certified staff. Response options for years at the school include: 0-1 year, 2-5 years, 6-10 years, and more than 10 years. The demographic item regarding respondents' involvement with the implementation team at their school was be recoded into two categories: "Not a member of the school's implementation team," and "Member of the school implementation team." The SIS datasets consisted of respondent level data, but included no unique identifiers for those respondents. All SIS item responses (1= Not True of Me, 2, 3= Somewhat True of Me, 4, and 5= Very True of Me) were used to create new variables representing means for the four essential element domains and a total mean for each respondent.

Description of Sample and Data

Two states whose departments of education were implementing different integrated academic and behavior tiered response-to-intervention models administered the School Implementation Scale (Appendix A) during the consecutive school years of 2012-2013 and 2013-2014. The first state was located in the Midwest and ranked close to 20th among all states in area and population. The Midwestern state served about 900,000 K-12 students and all districts/schools participated in the CW (Collaborative Work) initiative, then districts and/or schools and their administrations had the option to choose whether to implement either SW-PBS (School-Wide Positive Behavioral Supports) or PLC (Professional Learning Communities) in their buildings. The SIS, as administered in the Midwestern state consisted of a total of 37 items; the 31 original SIS items plus 6 items regarding other state initiatives being implemented: Collaborative Data Teams (3 items), effective teaching practices (2 items), and Common

Formative Assessments (1 item). All of the schools in the Midwestern state that administered the School Implementation Scale were identified by their state department of education as needing professional development for improving school-wide academic achievement. In addition to the CW initiative, the state focused on the general category of Response to Intervention (RTI), and included information on system change models which have been shown to be effective in improving student achievement, such as School-Wide Positive Behavior Supports (SW-PBS) and Professional Learning Communities (PLC). Professional learning communities see student learning, not teaching, as their mission, and focus on building the capacity of school personnel to create and sustain the conditions that promote high levels of student and adult learning. The SW-PBS three-tiered RTI model process creates safer and more effective schools by structuring the learning environment to support the academic and social success of all students.

The second state was located in the West, ranked in the top 10 among all states for area and close to 30th for population, and served about 500,000 students. All districts/schools participated in the EBISS (Effective Behavioral and Instructional Support Systems) initiative, then districts and/or schools and their administrations had the option to choose whether to implement either SWPBIS (Schoolwide Positive Behavioral Intervention Supports) or RTI (Response-To-Intervention) in their buildings. The SIS as administered in the Western state consisted of a total of 42 items; 28 of the 31 original SIS items plus 14 items regarding other state educational initiatives: tiered levels of academic and behavior supports (13 items), and student progress in meeting the Common Core State Standards (1 item). The Western state's general education initiative, EBISS, provided the organizational tools for districts to blend two specific evidence-based practices: the School-wide Reading Model (SWRM) and the School-wide Positive Behavioral Intervention Support (SWPBIS or PBIS) model. The EBISS Teaming

Framework provided a structure for data-based implementation teams at the district and school levels, with the primary goal of improving student outcomes. The EBISS framework then allowed districts to choose between Response-To-Intervention (RTI) and Positive Behavioral Interventions and Supports (PBIS) models in order to meet the academic and behavioral needs of all students.

The integrated academic and behavior response-to-intervention tiered support systems implemented in these two states represent two of the major types of models most adopted in education: the Response-To-Intervention Model or RTI, and the Positive Behavioral Supports Model which is implemented under various names such as School-Wide Positive Behavior Supports (SW-PBS) and Positive Behavioral Interventions and Supports (PBIS). RTI and PBIS are both implemented in the Western state and SW-PBS is implemented in the Midwestern state.

The scale was hosted on each of the participating states State Personnel Development Grant websites, which was also where responses were collected. The SIS was administered through a standardized process which included providing instructions to each school's principal with sample text for email distribution to staff (see Appendix B). Staff accessed and completed the SIS, which required approximately 5-10 minutes. Individual respondent data was kept confidential as the state, district, or schools received no individual level results; only overall summaries and reports aggregated by the demographic questions at the beginning of the survey (school, staff role, and membership on the implementation or leadership team) are disseminated. All data, including demographic information, questionnaire responses, and scores were kept on a password secured computer in a locked office to which only the researcher had access (see Appendix E). School summary reports provided tables displaying scores and means for each item and domain, and stacked bar graphs for the same data were available on the respective State

Personnel Development Grant websites. The SIS data from both states for 2012-2013 and 2013-2014 were downloaded from the websites in Excel files and converted into SPSS data files, as most of the data analysis was conducted with Statistical Package for the Social Sciences SPSS Version 24 (IBM Corp., 2016).

The 2012-2013 Midwestern state data consisted of 2,504 respondent records of the School Implementation Scale from 145 schools in 75 districts (see Table 5). The number of Midwestern respondents involved with SW-PBS was 706 from 34 schools in 25 districts. The 2012-2013 Western state data consisted of 1,236 respondent records of the School Implementation Scale from 110 schools in 24 districts. The number of Western respondents involved with SWPBIS was 437 from 34 schools in 8 districts, and the number of respondents involved with RTI was 237 from 17 schools in 4 districts (see Table 6).

Table 5. School Implementation Scale 2012-2013 State Data Frequencies and Descriptives

		Midwestern State CW		Western State EBISS	
		N	%	N	%
Staff Role	General Education Teacher	1741	69.5%	721	58.9%
	Special Education Teacher	276	11.0%	113	9.2%
	Administrator	80	3.2%	71	5.8%
	Other Certified Staff	348	13.9%	151	12.3%
	Non-Certified Staff	59	2.4%	168	13.7%
Years at School	0-1 year	255	10.2%	89	7.2%
	2-5 years	611	24.4%	214	17.3%
	6-10 years	565	22.6%	353	28.6%
	More than 10 years	1073	42.9%	580	46.9%
Implementation Team	Yes	1602	64.0%	439	37.2%
	No	902	36.0%	741	62.8%
TOTAL	Respondents	2504		1236	
	Districts	75		24	
	Schools	145		110	

Table 6. School Implementation Scale 2012-2013 Model Data Frequencies and Descriptives

		Midwestern State		Western State			
		SW-PBS		SWPBIS		RTI	
		N	%	N	%	N	%
Staff Role	General Education Teacher	491	69.5%	240	56.5%	161	67.9%
	Special Education Teacher	88	12.5%	35	8.2%	23	9.7%
	Administrator	19	2.7%	18	4.2%	14	5.9%
	Other Certified Staff	87	12.3%	64	15.1%	27	11.4%
	Non-Certified Staff	21	3.0%	68	16.0%	12	5.1%
Years at School	0-1 year	81	11.5%	38	8.5%	14	5.9%
	2-5 years	185	26.2%	79	18.1%	58	24.5%
	6-10 years	145	20.5%	110	25.2%	66	27.8%
	More than 10 years	295	41.8%	211	48.3%	99	41.8%
Implementation Team	Yes	496	70.3%	128	30.7%	97	40.9%
	No	210	29.7%	289	69.3%	140	59.1%
TOTAL	Respondents	706		437		237	
	Districts	25		8		4	
	Schools	34		34		17	

The Midwestern state data for 2013-2014 consisted of 3,129 SIS respondents from 206 schools in 84 districts (see Table 7). Seventy-four percent of respondents were on the implementation team, and 26% were not. General education teachers comprised 72% of Midwestern respondents, 12% were administrators, 11% other certified staff, 3% special education teachers, and 2% non-certified staff. Forty-two percent of the Midwestern respondents had been at their schools for more than 10 years, 24% for 6-10 years, 23% for 2-5 years, and 12% for one year. For 2013-2014, the Western state data included 1,375 SIS respondents from 130 schools in 31 districts. Thirty percent of respondents were on the implementation team, and 70% were not. General education teachers comprised 62% of Western respondents, 14% were other certified staff, 11% non-certified staff, 9% special education teachers, and 4% administrators. Forty-seven percent of the Western respondents had been at their schools for more than 10 years, 30% for 6-10 years, 13% for 2-5 years, and 11% for one year.

Table 7. 2013-2014 School Implementation Scale State Data Frequencies and Descriptives

		Midwestern State CW		Western State EBISS	
		N	%	N	%
Staff Role	General Education Teacher	2260	72.2%	851	61.9%
	Special Education Teacher	102	3.3%	124	9.0%
	Administrator	375	12.0%	60	4.4%
	Other Certified Staff	333	10.6%	193	14.0%
	Non-Certified Staff	59	1.9%	147	10.7%
Years at School	0-1 year	365	11.7%	145	10.5%
	2-5 years	723	23.1%	180	13.1%
	6-10 years	739	23.6%	407	29.6%
	More than 10 years	1302	41.6%	643	46.8%
Implementation Team	Yes	2323	74.2%	401	29.9%
	No	806	25.8%	940	70.1%
TOTAL	Respondents	3129		1375	
	Districts	84		31	
	Schools	206		130	

The number of Midwestern respondents involved with SW-PBS was 955 from 60 schools in 36 districts (see Table 8). Seventy-eight percent were involved with the implementation team, and 22% were not. General education teachers comprised 73% of the Midwest SW-PBS respondents, 12% were administrators, 10% other certified staff, 5% special education teachers, and 1% non-certified staff. Forty-one percent of the SW-PBS respondents had been at their schools for more than 10 years, 25% for 6-10 years, 22% for 2-5 years, and 12% for one year. The number of Western respondents involved with SWPBIS was 588 from 51 schools in 8 districts, and the number of respondents involved with RTI was 269 from 16 schools in 4 districts (see Table 8). Twenty-six percent of the Western SWPBIS respondents were involved with the implementation team, and 74% were not. General education teachers comprised 69% of the SWPBIS respondents, 12% were other certified staff, 9% non-certified staff, 8% special education teachers, and 3% administrators. Forty-seven percent of the Western SWPBIS had been at their schools for more than 10 years, 27% for 6-10 years, 13% for 2-5 years, and 13% for one year. Thirty-two percent of the Western RTI respondents were involved with the implementation team, and 68% were not. General education teachers comprised 63% of the RTI respondents, 15% were non-certified staff, 12% other certified staff, 5% special education teachers, and 5% administrators. Fifty percent of the Western RTI respondents had been at their schools for more than 10 years, 33% for 6-10 years, 7% for 2-5 years, and 10% for one year.

Table 8. 2013-2014 School Implementation Scale Model Data Frequencies and Descriptives

		Midwestern State		Western State			
		SW-PBS		SWPBIS		RTI	
		N	%	N	%	N	%
Staff Role	General Education Teacher	723	72.7%	406	69.0%	170	63.2%
	Special Education Teacher	46	4.6%	45	7.7%	13	4.8%
	Administrator	116	11.7%	15	2.6%	13	4.8%
	Other Certified Staff	102	10.3%	72	12.2%	33	12.3%
	Non-Certified Staff	8	0.8%	50	8.5%	40	14.9%
Years at School	0-1 year	116	11.7%	77	13.1%	28	10.4%
	2-5 years	220	22.1%	76	12.9%	18	6.7%
	6-10 years	252	25.3%	161	27.4%	89	33.1%
	More than 10 years	407	40.9%	274	46.6%	134	49.8%
Implementation Team	Yes	777	78.1%	155	26.4%	85	31.6%
	No	218	21.9%	433	73.6%	184	68.4%
TOTAL	Respondents	955		588		269	
	Districts	36		8		4	
	Schools	60		51		16	

The statistical data analyses conducted for the current study primarily concentrated on the data collected in 2013-2014 using IBM SPSS StatisticsVersion 24 (IBM Copr., 2016) and M*plus* Statistical Modeling Software Version 7.4 (Muthen & Muthen, 2015), a statistical modeling program. The 2012-2013 data was employed for cross-validation. The data analyses procedures were organized around the research questions. Groups to be analyzed and compared included the overall educational initiatives of Collaborative Work (CW) from the Midwestern state and

Effective Behavioral and Instructional Support Systems (EBISS) from the Western state; School-Wide Positive Behavior Supports (SW-PBS) from the Midwestern state and Schoolwide Positive Behavioral and Instructional Supports (SWPBIS) from the Western state; and Response-To-Intervention (RTI) from the Western state.

Research Question 1:

For each state's 2014 data, and for each of the different integrated academic and behavior tiered RTI models, how well does the four-factor essential elements of successful schools structure from the School Implementation Scale fit the data; what are the relationships among the factors of the scale?

The School Implementation Scale data from 2013-2014 was analyzed separately for each of the two states. The scale as administered in each state, with the additional items tailored to meet the needs of the state and their respective educational initiatives, was used for exploratory factor analysis (Tabachnick & Fidell, 2013) in SPSS. For the Midwestern state this included responses to the full 31-item School Implementation Scale, 3 items pertaining to Collaborative Data Teams, 2 items regarding effective teaching practices and 1 item about Common Formative Assessments; for a total of 37 items. For the Western state this included responses for 28 items from the School Implementation Scale, 14 items pertaining to tiered levels of academic and behavior supports, and 1 item regarding student progress in meeting the Common Core State Standards; for a total of 42 items. The data reduction technique of exploratory factor analysis (EFA) was conducted for each state's overall data separately in order to examine the structure of the scale as it was administered, and to examine whether the additional items had an impact on the factor structure of the School Implementation Scale. EFA was chosen for the analysis method due to the exploratory nature of extracting the number of latent constructs which were estimated including the additional items, not originally part of the School Implementation Scale. Principle axis factoring (Tabachnick & Fidell, 2013) was used as the extraction procedure, followed by the oblique rotation method Promax (Tabachnick & Fidell, 2013). There are a variety of extraction and rotation methods available to use during exploratory factor analysis; but each method is based upon different principles and is appropriate in specific circumstances, with certain types of data, and for different desired outcomes. One of the main criteria for choosing extraction and rotation methods is the orthogonality of the constructs or extent to which the factors are expected to correlate. The set of constructs represented by the SIS items, the essential elements of successful schools domains, have been shown to be correlated in the research literature (Cohen, McCabe, Michelli, & Pickeral, 2009; Kirk & Jones, 2004).

Principal Axis Factoring was employed as the method for the extraction of the factors in each exploratory factor analysis because it uses an iterative process to estimate the communalities, and its goal is to extract the maximum orthogonal variance from the dataset with each succeeding factor (Tabachnick & Fidell, 2007). Principal axis factor extraction is widely used and understood, and it "conforms to the factor analytic model in which common variance is analyzed with unique and error variance removed" (p.636). Rotation methods in factor analyses are categorized by whether they are orthogonal when extracted factors are not expected to be correlated, and oblique when the extracted factors may be correlated or are allowed to correlate. Assuming that there were correlations between the factors, as shown in the research literature for the essential elements of successful schools (Cohen, et al., 2009; Kirk & Jones, 2004), an oblique rotation method was used. The method of Promax Rotation with Kaiser Normalization was employed for the rotation procedures as it rotates orthogonal factors into oblique positions (Tabachnick & Fidell, 2007) to ease interpretability of the factor structure.

For each state's and each integrated multi-tiered academic and behavioral intervention model's factor structures for the essential elements domains, the Eigenvalues for each factor extracted, along with the individual and total percentages of variance accounted for by the factors, were examined. Rotated structure component item-factor loadings were also explored for both factor structures within each state.

The responses for the common 28 items of the state scales were used for confirmatory factor analyses in M*plus* and exploratory factor analysis SPSS, and were conducted for each state's overall integrated academic and behavior tiered support model (CW from the Midwestern state and EBISS from the Western state), and for the two different integrated academic and behavior tiered RTI models (SW-PBS from the Midwestern state and SWPBIS from the Western state, and RTI from the Western state).

Confirmatory factor analysis was used, as opposed to exploratory factor analysis (Tabachnick & Fidell, 2007), as there was already a hypothesized structure among the items based on theoretical research regarding the four essential elements of successful schools: school culture, evidence-based practices, on-going professional development, and family engagement. CFA is used to study the relationships between a set of observed variables and a set of continuous latent variables (Tabachnick & Fidell, 2013). Structural equation modeling (SEM) includes models in which regressions among the continuous latent variables are estimated. "SEM has two parts: a measurement model and a structural model. The measurement model for both CFA and SEM is a multivariate regression model that describes the relationships between a set of observed dependent variables and a set of continuous latent variables" (Muthen & Muthen, 2012, p. 55). The observed dependent variables are referred to as factor indicators and the continuous latent variables are referred to as factors, and their relationships are described by a set of linear regression equations. “The structural model describes three types of relationships in one set of multivariate regression equations: the relationships among factors, the relationships among observed variables, and the relationships between factors and observed variables that are not factor indicators” (Muthen & Muthen, 2012, p. 56). These relationships are also described by a set of linear regression equations employing maximum likelihood estimation procedures.

Rotation methods in factor analyses are categorized by whether they are orthogonal when extracted factors are not allowed or expected to correlate, and oblique when the extracted factors may be correlated or are allowed to correlate. The four SIS domains or essential elements of successful schools were presumably correlated, and therefore, the method of Promax Rotation with Kaiser Normalization was employed for the rotation procedures as it rotates orthogonal factors into oblique positions (Tabachnick & Fidell, 2007) to ease interpretability of the factor structure.

The CFA results of the rotated four-factor (essential elements domains) structure solutions from each of the two state’s overall models (Midwestern Collaborative Work and Western Effective Behavioral and Instructional Support Systems), and each of the specific RTI models (Midwestern SW-PBS and Western SWPBIS, and Western RTI) implemented in the states for the 2013-2014 data were examined according to their overall fit indices and the indictor or item-factor loadings. The overall fit indices explored for each of the four models fit to the data for were: the Chi-Square Test of Model Fit, the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SMSR). Guidelines for the interpretation of these overall fit indices were located in various sources. For the Chi-Square Test of Model Fit, a lower Chi-Square value with non-significant p-value indicates better fit (Barrett, 2007). For the CFI, values greater than .90 or .95

indicate reasonably good fit of the data to the model (Hu & Bentler, 1999). According to Browne & Cudeck (1993), an RMSEA value of less than or equal to .05 signifies a close approximate fit, values between .05 and .08 signify a reasonable error of approximation, and values greater than or equal to 1.0 signify poor fit of the structure model. Hu and Bentler (1999) and Steiger (2007) advocates for cut-off values of RMSEA at .06 and .07, respectively. For the SRMR, a value of less than .10 is considered favorable with .05-.08 signifying acceptable fit (Hu & Bentler, 1999).

Tabachnick & Fidell (2006) recommended that only variables or items with loadings of 0.32 or above should be interpreted, and stated that "the greater the loading, the more the variable is a pure measure of the factor" (p.649). Comrey & Lee (1992) provided rules of thumb for interpreting factor loadings. They suggested that factor loadings higher than 0.71 are considered excellent, 0.63 very good, 0.55 good, 0.45 fair, and 0.32 poor.

The relationships between the latent variables derived from the four-factor structure solutions for the SIS were examined through Pearson Product-Moment Correlations (Pearson, 1901). The correlations among the essential elements of successful schools domains factor scores (school culture, evidence-based practices, on-going professional development, and family engagement) were computed and their significance explored. Additionally, these correlations and significance values were examined across states' overall tiered RTI models of CW and EBISS, and across the different specific tiered support models of Midwestern SW-PBS and Western SWPBIS, and Western RTI.

Research Question 2:

For the 2014 data (cross validated with 2013 data), to what extent does measurement invariance for the latent constructs measured by the School Implementation Scale, its four essential element domains, hold over states with different overall integrated academic and *behavior tiered response-to-intervention models; specifically, Collaborative Work (CW) from the Midwestern state and Effective Behavioral and Instructional Support Systems (EBISS) from the Western state?*

Measurement invariance (or factorial invariance) concerns the extent to which the psychometric properties of the observed indicators of a measurement instrument are generalizable. At its most basic level, factor invariance is whether the factors in each group of participants are measuring the same construct in the same way in different groups or over time. The observed scores of a measure should depend only on the latent construct scores, not on group membership or occasion. They should reflect only the true differences between groups in the amount or variability of the construct being measured. To be valid for group comparisons an instrument should be invariant, or measure identical constructs within the same structure across different groups (van de Schoot, Lugtig, & Hox, 2012).

Measurement or factorial invariance would have been tested in a latent trait modeling framework, specifically within the multiple group confirmatory factor model (Vandenberg & Lance, 2000), across the Collaborative Work (CW) group from the Midwestern state and Effective Behavioral and Instructional Support Systems (EBISS) group from the Western state. The process of testing factorial invariance has two distinct parts: measurement invariance (whether or not the construct is being measured in same way across groups) and structural invariance (whether or not groups differ in their distribution and/or means of the construct). Measurement model invariance consists of items or indicators having the same factor loadings, the same intercepts, and possibly the same residual covariance. Structural model invariance consists of indicators having the same factor variances, covariances, and means.

Vandenberg & Lance (2000) stated that "if not tested, violations of measurement equivalence assumptions are as threatening to substantive interpretations as is an inability to demonstrate reliability and validity" (p.6), but measurement equivalence is readily testable within a multiple group confirmatory factor analysis framework. A series of tests used to establish measurement invariance across groups from the bottom up was outlined by Vandenberg & Lance in 2000, and their sequence of steps would have been conducted for this study using Mplus Statistical Modeling Software Version 7.4 (Muthen & Muthen, 2015) for latent variables. Mplus Version 7.11 added a convenient shortcut for conducting a series of increasingly restrictive invariance tests (Bowen, 2014), and the theta parameterization approach to the specification of the multiple group confirmatory factor analysis would have been employed as it allows the researcher to obtain information on residual variances or the unexplained variance in the observed indicators of factors (Muthén, & Muthén, 2012). The appropriate variance-covariance matrices resulting from the groups of respondent data would have been created and read into M*plus* for analysis. Maximum likelihood estimation was planned, and change in model fit would have been assessed by implementing the likelihood ratio test in model χ^2. The planned statistical analysis consisted of the following steps.

Step 1: Test of the difference of the variance-covariance matrices relating items in the same measure across the groups being compared. This omnibus test compares the full saturated covariance matrix for both groups. If the p-value of the χ^2 with its associated degrees of freedom is significant, the data fail the omnibus test and further investigation of the invariance of the groups is necessary. If the test indicates lack of difference the usual conclusion is that measures are invariant (the matrices do not differ between groups on the whole) and no further tests are employed.

Step 2: Test of configural invariance which requires a demonstration that the same factors and pattern of factor loadings explains the variance-covariance matrices associated with the groups' responses, which means that the factor structure implied is the same for two or more groups of respondents (although values of parameters may vary). Separate analyses of subgroup variance-covariance matrices are conducted to determine if a common model in which all model parameters are allowed to differ between groups can reasonably represent both matrices and be used as the baseline against which other more restrictive models of the data are compared. Thus, the model χ^2 and its associated degrees of freedom would be additive across groups.

Step 3: Test of metric equivalence which examines whether the values of the factor loadings of each item on each factor are the same across groups. Items are allowed to have different loadings, but loadings for the same items are constrained to equality across groups. The fit of the metric invariance model is compared to the configural model, and if the p-value for the χ^2 statistic with its associated degrees of freedom is significant, then the data fail the metric invariance test and not all items have invariant factor loadings across groups. Demonstration of the equivalence of factor loadings is labeled "strong invariance," whereas configural invariance is sometimes called "weak invariance" (Horn & McArdle, 1992). If the fit of the metric model is not significantly different than the configural model, the modification indices are inspected to see if there are any items whose loadings differ between groups. The modification indices can be different between groups or on specific items. The model is re-tested, but because modification indices are single degree of freedom tests, and items must be added/parameter constraints released for one item at a time until the items with different loadings are identified. If at least one loading on a factor differs, this indicates partial metric invariance in which the overall construct is determined to be measured in same way across groups. If there are no non-invariant items identified, further tests are not necessary.

Step 4: Test for scalar invariance which requires that the intercepts of the regression equations of the observed variables on the latent factors are equivalent across groups. This step investigates whether groups have the same item intercepts. The scalar invariance model says that factor mean differences can cause item mean differences, however, the item intercepts should still be the same. For analysis each item is allowed to have a different intercept, but the intercepts for the same item across groups are constrained to equality. To estimate the intercepts, the factor mean is constrained to 0 in the reference group, and only those intercepts for which metric invariance holds are tested. The fit of scalar invariance model is compared to the metric invariance model, and if not significant then all modification indices are inspected to see if there are any items whose intercepts differ between groups. The model is be retested as needed; releasing one item at a time, to identify the items with differing intercepts between groups. If at least one item is deemed to have a non-invariant intercept, it indicates partial scalar invariance.

Step 5: Test of the equality of the residual variance associated with each observed variable which means that the residuals of the regression equation for each indicator are equivalent across groups. This is also called strict factorial invariance; where each item is still allowed to have different residual variances, but across groups each item's residual variance is constrained to equality. Testing residual variance is the last step in assessing measurement invariance, and only those items with differing residual variances and for which metric and scalar invariance already hold are tested. The fit of residual invariance model is compared to the scalar invariance model. If not significant, the modification indices are inspected to identify any items whose residual variances differ between groups - indicating partial residual variance.

The first five steps in the process are described by Schmitt & Kulijanin (2008, p. 212) as the process of addressing issues of measurement invariance, and the subsequent steps for the planned statistical analysis involved investigating the relationships between the latent factors themselves, which are often referred to as issues related to structural invariance. If measurement invariance holds, it is not problematic for structural invariance not to hold because it becomes a substantive issue about differences in the latent trait amounts and relations.

Step 6: Test whether factor variances are equal across groups while the variances of the latent factors are constrained to equality across groups. The factor variance is fixed to 1 for both groups, and if model fit becomes worse, then groups differ in their factor variances. **Step 7:** Test whether factor variances are equal across groups while the covariances of the latent factors are further constrained to be equal. The factor covariances are fixed to equality across groups, and factor correlations examined. The factor correlations are the same across groups if the factor variances are equal as well. **Step 8:** Test of the equivalence of factor means in which one mean is set to zero and the significance of the other parameter represents a test of the difference of latent means (Schmitt & Kulijanin, 2008). The factor means in both groups are fixed at 0, and if model fit gets worse, then groups differ in their factor means.

The series of statistical tests for measurement invariance as delineated by Vandenberg and Lance (2000) were planned for the School Implementation Scales' four-factor essential elements of successful schools domains structure across the Midwestern state's Collaborative Work (CW) Initiative and the Western state's Effective Behavioral and Instructional Support Systems (EBISS) overall. Cross validation with the 2012-2013 data was also planned if measurement invariance had not held.

Chen (2007) provided recommendations regarding the cutoff points based on the three routinely used model fit indices (i.e., CFI, RMSEA, and SRMR) for evaluating invariance at the three commonly tested levels, as derived from the results of two simulation studies. Similar values are suggested for CFI and RMSEA across all three levels of invariance tests, but different values are proposed for SRMR, as SRMR is more sensitive to non-invariance in loadings than to non-invariance in intercepts or residual variances (Chen, 2007). When the sample size is small ($N < 300$), sample sizes are unequal, and the pattern of non-invariance is uniform, the following cutoff criteria were suggested:

- For testing loading invariance, a change of greater than or equal to -.005 in CFI, supplemented by a change of greater than or equal to .010 in RMSEA or a change of greater than or equal to .025 in SRMR, would indicate non-invariance.
- For testing intercept or residual invariance, a change of greater than or equal to .005 in CFI, supplemented by a change of greater than or equal to .010 in RMSEA, or a change of greater than or equal to .005 in SRMR, would indicate non-invariance.

When sample size is adequate (total $N > 300$) and sample sizes are equal across the groups, particularly when lack of invariance is mixed, the following more stringent criteria were suggested:

- For testing loading invariance, a change of greater than or equal to -.010 in CFI, supplemented by a change of greater than or equal to .015 in RMSEA, or a change of greater than or equal to .030 in SRMR, would indicate non-invariance.
- For testing intercept or residual invariance, a change of greater than or equal to -.010 in CFI, supplemented by a change of greater than or equal to .015 in RMSEA, or a change of greater than or equal to .010 in SRMR, would indicate non-invariance.

Among the three indices, CFI was chosen by Chen (2007) as the main criterion because RMSEA and SRMR tend to over-reject an invariant model when sample size is small and because changes in RMSEA are more likely to be affected by sample size and model complexity. “However,” stated Chen (2007, p.502), “these criteria should be used with caution, because testing measurement invariance is a very complex issue. As uncovered in this investigation, a number of factors can affect the magnitude of changes in fit statistics, such as pattern of noninvariance, sample size, ratio of sample size, and model complexity.”

Research Question 3:

For the 2014 data (cross validated with 2013 data), to what extent does measurement invariance for the latent constructs measured by the School Implementation Scale, its four essential element domains, hold over different major types of integrated academic and behavior tiered response-to-intervention models within and between states; specifically Response-to-Intervention (RTI) from the Western state, and School-Wide Positive Behavior Supports (SW-PBS) from the Midwestern state and School-Wide Positive Behavior Interventions and Supports (SWPBIS) from the Western state?

The steps for testing measurement invariance as outlined by Vandenberg and Lance (2000) would have been repeated to analyze the extent to which measurement invariance held across specific types of integrated academic and behavior tiered models. The groups of School-Wide Positive Behavior Supports (SW-PBS) from the Midwestern state and School-Wide Positive Behavior Interventions and Supports (SWPBIS) from the Western state were planned to be compared with the Response-to-Intervention (RTI) group from the Western state. Cross validation with the 2012-2013 data was also planned if measurement invariance had not held.

Had the measurement invariance of the School Implementation Scale held between states or between groups to some level, research questions numbers 4 and 5 would have been conducted to the extent possible based on the results of measurement invariance testing.

Research Question 4:

For the 2014 data, does the School Implementation Scale instrument overall and its four essential elements of effective schools domains provide reliable data across the state samples (Midwestern CW and Western EBISS) and across the different major types of integrated academic and behavior tiered response-to-intervention models (Midwestern SW-PBS and Western SWPBIS, and Western RTI) for all school staff in terms of respondent's role, years of experience, and involvement with the leadership team?

A reliable survey is free from measurement error generated from poorly worded questions, poor design, ambiguous terms, inappropriate reading level, or unclear directions (DeVellis, 2003). Cronbach's coefficient alpha (Cronbach, 1951) is most commonly used to evaluate reliability because it can be used to analyze a single form survey and can be applied to ordinal data (Gliem & Gliem, 2003). This procedure would have been conducted to estimate the consistency of responses across items and subsets of items for the School Implementation Scale. It was planned to analyze the 2013-2014 SIS item response data separately for each state. SIS item responses would have been used to calculate Cronbach's coefficient alpha values for the SIS overall, for each of the four essential element domains, and for each of the three levels of implementation subscales. These calculations of coefficient alpha for the overall scale and subscales were to be conducted using IBM SPSS Version 24 (IBM Corp., 2016) for each state overall and each of the integrated academic and behavior RTI models implemented by the two states in the dataset. These same Cronbach's alpha reliability estimates would have been calculated for each category of respondent staff role, years at the school categories, and implementation team membership. All Cronbach's alpha coefficients by staff role, years at the school, and implementation team membership were to be examined for patterns across states and between the states' different integrated academic and behavior multi-tiered RTI models being implemented.

Research Question 5:

For the 2014 School Implementation Scale results, what are the differences between and patterns observed in respondents' mean SIS and mean essential element domain results in terms of staff role, years at the school, involvement with the leadership team, and the interactions of these variables for each state overall (Midwestern CW and Western EBISS) and for each states' major types of integrated academic and behavior tiered response-to-intervention models (Midwestern SW-PBS and Western SWPBIS, and Western RTI)?

Statistical analysis of the 2013-2014 SIS respondent means for each state separately was planned, using IBM SPSS Version 24 (IBM Corp., 2016). Data would have been analyzed for group differences and patterns among groups using a series of factorial Multivariate Analyses of Variance (MANOVA) followed by simple contrast tests for the different levels of all IVs and pairwise post-hoc comparisons between the levels of each interaction of IVs (Tabachnick & Fidell, 2006) using the Bonferroni adjustment for alpha (Bland & Altman, 1995). The respondent means for the overall SIS and for each of the four essential elements of successful schools domains would have been employed as the dependent variables for the MANOVA. These differences and patterns across the two states and their integrated academic and behavior tiered RTI models were to be examined.

Multivariate Analysis of Variance (MANOVA) is a generalization of the Analysis of Variance (ANOVA) procedures to a situation in which there are more than one dependent variable. ANOVA procedures allow researchers to assess the relationship of one or more factors with one DV (Green & Salkind, 2011). "Factors are either between-subjects or within-subjects factors. A between-subjects factor divides research participants into different groups such as gender or multiple treatments. A within-subjects factor has multiple levels, and each participant is observed on a dependent variable across those levels" (Green & Salkind, 2011, p. 182). The factor divides individuals into two or more groups or levels, while the dependent variable differentiates individuals on a quantitative dimension, and the ANOVA F-test evaluates whether group means on the dependent variable differ significantly from each other. MANOVA examines whether a combination of measures vary as a function of the groups' characteristics (Tabachnick & Fidell, 2006). If a main effect for an IV with more than two levels or categories are significant, simple contrast comparisons should be conducted to examine which levels of main effects are significantly different from one another. If an interaction is significant, follow-up pairwise comparison of cells in the interaction should be conducted to examine which exact groups are different from which other groups and what patterns or trends can be identified. The Bonferroni adjustment for alpha levels was to be employed for the follow-up analyses (Bland & Altman, 1995).

Independent variables for the planned analyses would have included respondents' staff role types (administrators, general education teachers, special education teachers, other certified staff, or non-certified staff), respondents' years at the school ("0-1 years," "2-5 years," "6-10 years," or "More than 10 years"), respondents' involvement with the school implementation team (member of the implementation team or not), and the potential interactions of these variables. It was hypothesized that means of respondents involved with the implementation team would be higher than those of respondents not involved with the implementation team (Walker, 2006; Gaumer Erickson, et al., 2012). It was also hypothesized that administrators would have higher means than general education or special education teachers, whose means would be higher than other certified or non-certified staff. In terms of years at the school, it was hypothesized that those respondents who had "0-1 year" and "more than 10 years" would have lower means than their colleagues who have "2-5 years" or "6-10 years." There was a hypothesized possibility that two-way interactions between role and years at the school, role and implementation team involvement, or years at the school and implementation team involvement could produce anomalies in these expectations. Further, a possible three-way interaction among the respondent characteristics of role, years at the school, and implementation team involvement could have also lead to different outcomes in group SIS means than those hypothesized.

It was planned to analyze the main effects or mean differences in the composite DV among groups of respondents by staff role, years at the school and involvement with the implementation team at different levels of each IV holding all else constant. The marginal means are the best estimates of population parameters for main effects, and cell means are the best estimates of population parameters for interactions (Tabachnick & Fidell, 2006). Two-way interactions between the levels of staff role and years at school, staff roles and implementation team involvement, years at school and implementation team involvement, and the three-way interaction of staff role, years at the school, and implementation team membership would have also been analyzed. Significant differences for one or more of the main effects or interactions could reveal which DVs are changed and which are unaffected by the IVs (Tabachnick & Fidell, 2006). If a main effect for an IV with more than two levels or categories were significant, simple contrast comparisons would have been conducted to examine which levels of main effects were significantly different from one another. If an interaction were significant, follow-up pairwise comparison of cells in the interaction would have been conducted to examine which exact groups were different from which other groups and what patterns or trends could have been identified.

For the multivariate tests, Pillai's Trace and Wilke's Lambda F statistics with their associated degrees of freedom and p-values for significance were to be investigated. Partial Eta Squared effect sizes would have also been provided for all multivariate results. For the univariate contrast results of each IV the F statistics with their associated degrees of freedom, p-values or significance levels, and the Partial Eta Squared effect sizes would have been examined. For the follow-up multiple comparisons, mean differences along with significance values and confidence intervals would have been investigated. The estimated marginal means and cell means from the pairwise comparisons would have assisted in the interpretation of the MANOVA results as to how respondent characteristics impact the SIS results. The results of the group comparisons for each state and each their integrated academic and behavior RTI model were to be compared to investigate the how the SIS results are impacted by various respondent characteristics.

Results

The exploratory and confirmatory factor analysis conducted in SPSS and Mplus, respectively, on the five datasets that consisted of the common 28 School Implementation Scale items (Midwestern CW, Western EBISS, SW-PBS/SWPBIS, RTI, and both states combined) revealed that the 4-factor structure specified for the SIS does not adequately fit any of the datasets examined. However, the SIS originally included 31 items. Statistical analysis conducted on the 2014 Midwestern state's CW dataset consisting of the original 31 SIS items found that the data did not adequately fit the 4-factor structure specified for the SIS. This finding was cross-validated with the 2013 Midwestern CW 31-item data which did not adequately fit the model. Therefore, the 4-factor structure may be an incorrectly specified model for the School Implementation Scale, and should be reviewed. This study concludes that the SIS with its current 4-factor structure, for either 28 or 31 items, is not a valid and reliable measure of the implementation level of integrated academic and behavioral multi-tiered systems, and should not be used across various state populations and different forms of multi-tiered support systems.

Preliminary Analysis

Missing Data. The 2013-2014 Midwestern state data files included no missing data or such data had already been deleted list-wise before received by the researcher. The 2013-2014 Western state dataset did have missing data among the 1,375 respondents which was analyzed and found to be missing completely at random (MCAR results). Of all 42 items included on the Western state's administered version of the School Implementation Scale, individual items revealed only 5-37 missing values (0.36%-2.69%) across the full state dataset, and no patterns were detected among the missing values even though 33 cases were found to contain more than 50% of item responses missing. For the Western Response-to-Intervention data, which included 269 respondents, only 1-9 values (0.3%-3.35%) were missing for each of the 42 items. For the Western Schoolwide Positive Behavior Intervention Supports data, which included 588 respondents, 2-17 values (0.34%-2.90%) for each of the 42 items were missing. No patterns among the missing data were detected for either the Western state's RTI or SWPBIS respondent data. Missing values were dealt with using pair-wise deletion during analysis instead of list-wise deletion to lessen the impact on the number of cases per item available for analysis.

Assumptions. Univariate assumptions of normality and multicollinearity were examined. Frequency distributions and means compared to modes for almost all items showed negative skewness from the normal distribution, and large standard deviations of the means showed levels of kurtosis present for almost all items. For the Midwestern Collaborative Work 2013-2014 data, item means for the 28 common School Implementation Scale items ranged from 3.23 to 4.54. Sixteen of the 28 items (57%) had median scores that were greater than their means, and 26 of the 28 items (93%) had modes that were greater than their means. Skewness values for the items ranged from -1.9 to -0.217 with a standard error of skewness equal to 0.044. The standard deviation for the items ranged from 0.658 to 1.301, and 9 of the 28 items (32%) had standard errors of the mean greater than 1.0. Kurtosis values for the items ranged from -0.884 to 4.169 with a standard error of kurtosis equal to 0.088. The additional 7 items included on the School Implementation Scale for the Midwestern CW data at the request of the state also displayed high means with even higher medians and modes, and large standard deviations of the means.

For the Western Effective Behavioral and Instructional Support Systems 2013-2014 data, item means for the 28 common School Implementation Scale items ranged from 2.58 to 4.16. Twenty-two of the 28 items (79%) had median scores that were greater than their means, and 25 of the 28 items (89%) had modes that were greater than their means. Skewness values for the items ranged from -1.33 to 0.32 with a standard error of skewness equal to 0.066. The standard deviation for the items ranged from 0.949 to 1.367, and 22 of the 28 items (79%) had standard errors of the mean greater than 1.0. Kurtosis values for the items ranged from -0.944 to 1.166 with a standard error of kurtosis equal to 0.926. The additional 14 items included on the School Implementation Scale for the Western EBISS data at the request of the state also displayed moderately high item means with approximately equal or higher medians and modes, as well as large standard deviations of the item means.

For both the Midwestern CW and Western EBISS 2013-2014 data, items means for the 28 common School Implementation Scale items ranged from 3.03 to 4.41. Sixteen of the 28 items (57%) had median scores that were greater than their means, and 26 of the 28 items (93%) had modes that were greater than their means. Skewness values for the items ranged from -1.776 to -0.064 with a standard error of skewness equal to 0.036. The standard deviation for the items ranged from 0.812 to 1.321, and 16 of the 28 items (57%) had standard errors of the mean greater than 1.0. Kurtosis values for the items ranged from -1.022 to 3.634 with a standard error of kurtosis equal to 0.073. The 2013-2014 School Implementation Scale item data for the Midwestern state CW initiative, the Western state's EBISS initiative, or both states' data combined does not seem to be normally distributed.

Multicollinearity was checked by conducting all pairs of bivariate correlations in each of the data sets, and correlation tables can be found in Appendix C. For the Midwestern CW data, all item correlations were significant at the $p<0.01$ level, and two correlations between items approached a value larger than 0.800, indicating multicollinearity (Tabachnick & Fidell, 2013). A significant Pearson correlation of 0.870 revealed that items F2 (I regularly communicate with families regarding student academic goals/progress) and F3 (I regularly communicate with families regarding student behavioral goals/progress) were essentially measuring the same construct. Another significant Pearson correlation of 0.736 revealed that item E12 (When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions) and item E13 (When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions) were approximately measuring the same construct.

For the Western EBISS data, all but 3 item correlations were significant at the $p<0.05$ or $p<0.01$ levels, and two of the correlations approached multicollinearity. The non-significant bivariate Pearson correlations included item C5 (I have the technology and resources that I need to provide effective instruction) and F3 (I regularly communicate with families regarding student behavioral goals/progress) with $r=0.028$ and $p=0.305$; item C5 (I have the technology and resources that I need to provide effective instruction) and F4 (I make informed decisions based on feedback from families) with $r=0.043$ and $p=0.107$; and item C5 (I have the technology and resources that I need to provide effective instruction) and E4 (I review universal screening data at least three times a year for every student that I support) with $r=0.051$ and $p=0.060$. A significant Pearson correlation of 0.798 showed that items F3 (I regularly communicate with families regarding student behavioral goals/progress) and F4 (I make informed decisions based on feedback from families) were very close measuring the same construct. Another significant Pearson correlation of 0.748 showed that items F2 (I regularly communicate with families regarding student academic goals/progress) and F3 (I regularly communicate with families regarding student behavioral goals/progress) were also close measuring the same construct.

For both the Midwestern CW and Western EBISS 2013-2014 dataset together, all item correlations were significant at the $p<0.01$ level, and several correlations approached levels of multicollinearity (Tabachnick & Fidell. 2006). A significant Pearson correlation of 0.833 showed that items F2 (I regularly communicate with families regarding student academic goals/progress) and F3 (I regularly communicate with families regarding student behavioral goals/progress) were measuring the same construct. Another significant Pearson correlation of 0.757 between items E12 (When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions) and E13 (When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions), and a significant correlation of 0.707 between items E7 (I adapt the environment, curriculum, and instruction based on my students' academic data) and E8 (I adapt the environment, curriculum, and instruction based on my students' behavioral data) showed that these pairs of items were approximately measuring the same construct.

Multivariate assumptions were also examined within the three data sets, and no extreme multivariate outliers were detected. The Midwestern CW data revealed a Box's M value of 16990.108 with $F(7308, 331016.488) = 1.936$, the Western EBISS data revealed a Box' M value of 5710.058 with $F(3654, 235716.055) = 1.297$, and both states' data combined revealed a Box's M value of 30142.242 with $F(12586, 606902.394) = 2.005$; and all of these values were significant at the $p<0.001$ level.

Multivariate tests for response trends within demographic categories of respondents showed non-significant results for several groups within the three datasets, meaning that there could be patterns for some groups on items or groups of items. The Midwestern CW data revealed a Pillai's Trace value of 0.033 with $F(84, 9160.940) = 1.217$, $p=0.087$ and a Wilk's Lambda value of 0.967 with $F(84, 9192) = 1.217$, $p=0.087$ for number of years the individual respondent had been employed. The combination of the number of years an individual had been employed with whether or not they were a member of the implementation team revealed a Pillai's Trace value of 0.028 with $F(84, 9192)=1.036$, $p=0.389$ and a Wilk's Lambda value of 0.972 with $F(84, 9160.94)=1.037$, $p=0.388$.

The Western EBISS data revealed a Pillai's Trace value of 0.099 with $F(112, 5240) = 1.186$, $p=0.0909$ and a Wilk's Lambda value of 0.904 with $F(112, 5193.466) = 1.189$, $p=0.087$ for role- whether the individual respondent was an administrator, teacher, other certified staff, or non-certified staff. The number of years an individual respondent had been employed displayed a Pillai's Trace value of 0.058 with F(84, 3927)= 0.928, p=0.663 and a Wilk's Lambda value of 0.943 with F(84, 3910.808)=0.664, p=0.664. The combination of role and years employed revealed a Pillai's Trace value of 0.232 with F(336, 15816)= 0.927, p=0.824 and a Wilk's Lambda value of 0.790 with F(336, 14492.207)= 0.927, p=0.824. The combination of role and whether or not the individual respondent was on the implementation team displayed a Pillai's Trace value of 0.086 with F(112, 5240)= 1.027, p=0.406 and a Wilk's Lambda value of 0.917 with F(112, 5193.466)= 1.028, p=0.403; the combination of years employed and whether or not the individual respondent was a member of the implementation team revealed a Pillai's Trace value of 0.060 with F(84, 3927)= 0.961, p=0.581 and a Wilk's Lambda value of 0.941 with F(84, 3910.808)= 0.961, p=0.580; and the combination of role, years employed, and membership on the implementation team displayed a Pillai's Trace value of 0.220 with F(336, 15816)= 0.878, p=0.947 and a Wilk's Lambda value of 0.800 with F(336, 14492.207)= 0.878, p=0.947. These values from the multivariate assumptions tests showed that categories of individual respondents and various combinations of those categories could have produced trends or patterns within their item responses on the SIS.

However, for the combined Midwestern CW and Western EBISS data, none of the respondent categories or combinations of categories produced non-significant Pillai's Trace or Wilk's Lambda values. This indicated that within the larger dataset including both states the individual respondent categories such a role, years of employment, membership on the implementation team, or any of the two-way or the three-way combination of those categories did not seem to produce any detectable significant trends or patterns among the item responses.

Levene's test of equality of error variances was conducted on all three datasets to examine whether the error variances of the items were equal across groups of respondents. The Midwestern CW data displayed significant F-values for all 28 of the common School Implementation Scale items indicating that there were not differences in the error variances of items by respondent groups. The Western EBISS data revealed 5 items with non-significant Levene's test results. Item C4 (I think my school has an effective process in place to identify available resources) resulted in $F(39, 1334)= 1.012$ with $p=0.451$, item C5 (I have the technology and resources that I need to provide effective instruction) resulted in $F(39, 1334)= 1.151$ with $p=0.243$, item C6 (I have the time necessary to analyze students data and problem-solve with my colleagues) resulted in $F(39, 1334)= 0.790$ with $p=0.819$, item E13 (When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions) resulted in $F(39, 1334)=1.374$ with $p=0.064$, and item F5 (I think my school does a good job of including parents as team members in data-based decision making) resulted in $F(39, 1334)= 0.836$ with $p=0.754$. These items seem to have error variances that were not equal across categories or groups or respondents for the Western EBISS data. Levene's test of equality of error variances revealed for the combined Midwestern CW and Western EBISS data, only item C6 (I have the time necessary to analyze students data and problem-solve with my colleagues) resulted in a non-significant $F(39, 4463)= 1.260$ with $p=0.129$, meaning that the error variances for the item were not equal across respondent category or groups.

Research Question 1: *For each state's 2014 data, and for each of the different integrated academic and behavior tiered RTI models, how well does the four-factor essential elements of successful schools structure from the School Implementation Scale fit the data; what are the relationships among the factors of the scale?*

Confirmatory Factor Analysis (CFA) was conducted in Mplus for each state's common 28 SIS items data, for each different tiered system of support model (SW-PBS/SWPBIS and RTI), and then for the dataset consisting of both states' common 28 SIS items. Mplus was employed for CFA because it provides model fit statistics to test how well the datasets fit the 4-factor structure theorized for the School Implementation Scale.

Exploratory Factor Analysis (EFA) specifying 4 factors was conducted in SPSS on the 28 common School Implementation Scale items for the Midwestern CW data; the Western EBISS data; the Midwestern SW-PBS and Western SWPBIS data; and the Western RTI data. After examining both Principal Components Analysis and Principal Axis Factoring as factor extraction methods and experimenting with different combinations of rotation methods (direct oblimin and Promax), it was decided that PAF factor extraction and Promax rotation for oblique factors provided the most interpretable exploratory factor structures for the data. Then EFA specifying 4 factors was conducted using SPSS on the 28 common SIS items from each state's datasets combined. Principle Axis Factoring was employed as the factor extraction method with Promax rotation for oblique factors, and a 4-factor solution was specified.

Exploratory Factor Analysis (EFA), without specifying the number of factors to extract and onto which the items should load, was conducted in SPSS on the separate states' data for only the 28 common items of the original 31 items of the School Implementation Scale; for the different tiered support system models of School Wide Positive Behavior Supports from the Midwestern state and Schoolwide Positive Behavior Intervention Supports from the Western state, the Response-to-Intervention from the Western state; and the combined 28 common SIS item data from both states. EFA without specifying the number of factors was also conducted in

SPSS separately for both the Midwestern CW and Western EBISS datasets from the School Implementation Scale as administered with 37 and 42 items, respectively.

Confirmatory Factor Analysis with Mplus

Midwestern State Collaborative Work 28 items- Confirmatory Factor Analysis was conducted on the 2014 Midwestern Collaborative Work state-wide initiative data in Mplus, consisted of the common 28 SIS items, and specified the 4-factor structure model outlined for the School Implementation Scale. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, Stage, Barlow, & King, 2006). The Comparative Fit Index (CFI) was equal to 0.795, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.775, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.094, but a value of 0.06 or less means acceptable model fit (Steigler, 2007). CFA results for the Midwestern state data rendered a significant Chi Square test of base model fit value of 47096.620 with 378 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value relative to the degrees of freedom are considered acceptable for good model fit (Hooper, Coughlin, & Mullen, 2008).

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA on the Midwestern CW 28-item data can be found in Appendix D Tables 1, 2, and 3. The standardized parameter estimates, or item-factor loading coefficients, for Factor #1 revealed a range from 0.338 to 0.829. The items with the lowest coefficients were C3, C8, C1, and C5; but items C2, C6, C7, C4, and C9 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.374 to 0.797. Items E1, E13, E2, E7, E8, and E12 showed standardized parameter estimates lower than 0.500; and items E4, E6, E11, and E3 had item-factor loadings larger than 0.600. Factor #3 revealed item-factor loadings ranging from 0.514 to 0.809, and while item P2 had a coefficient value around 0.500, the other 4 items had coefficient values larger than 0.700. For Factor #4 the standardized parameter estimates ranged from 0.364 to 0.870. Items F1 and F5 showed item-factor loading values less than 0.600; but items F3, F2, and F4 showed item-loading values larger than 0.600.

Factor score coefficients for the Midwestern CW 28-item CFA in Mplus ranged from 0.039 to 0.094 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 2). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.025 to 0.062 on Factor #2, and the Professional Development items revealed values ranging from 0.112 to 0.232 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.014 to 0.164 for the theorized Family Involvement items.

<u>Western State Effective Behavioral and Instructional Support Systems with 28 Items-</u>

Confirmatory Factor Analysis was conducted in Mplus on the 2014 Western Effective Behavioral and Instructional Support System state-wide initiative data which consisted of the common 28 SIS items, and specified the 4-factor structure model outlined for the School Implementation Scale. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, et al., 2006). The Comparative Fit Index (CFI) was equal to 0.766, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.743, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.102, but a value of 0.07 or less means acceptable model fit (Steigler, 2007). CFA results for the Western state data rendered a significant Chi Square test of base model fit value of 21370.636

with 378 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value relative to the degrees of freedom are considered acceptable for good model fit (Hooper, et al., 2008).

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA on the Western EBISS 28-item data can be found in Appendix D Tables 4, 5, and 6. The standardized parameter estimates, or item-factor loading coefficients, for Factor #1 revealed a range from 0.430 to 0.828. The items with the lowest coefficients were C3, C5, C1, and C6; but items C7, C2, C9, C4, and C8 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.629 to 0.947. Items E6, E13, E1, E8, and E7 showed standardized parameter estimates lower than 0.750; and items E4, E3, E12, E11, and E2 had item-factor loadings larger than 0.750. Factor #3 showed item-factor loadings ranging from 0.754 to 0.976, and while items P2 and P4 had coefficient values less than 0.900, items P3 and P1 had coefficient values larger than 0.900. For Factor #4 the standardized parameter estimates ranged from 0.280 to 1.141. Items F5 and F1 showed item-factor loading values less than 0.600; but items F3, F4, and F2 showed item-loading values larger than 0.900.

Factor score coefficients for the Western EBISS 28-item CFA in Mplus ranged from 0.027 to 0.113 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 5). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.046 to 0.094 on Factor #2, and the Professional Development items revealed values ranging from 0.103 to 0.177 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.011 to 0.178 for the theorized Family Involvement items.

School-Wide Positive Behavior Support & Schoolwide Positive Behavior Intervention Supports with 28 Items- Confirmatory Factor Analysis was conducted in Mplus on the 2014 School-Wide Positive Behavior Supports and Schoolwide Positive Behavioral Interventions and Supports data from both the Midwestern and the Western states combined, consisted of the common 28 SIS items, and specified the 4-factor structure model outlined for the School Implementation Scale. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, et al., 2006). The Comparative Fit Index (CFI) was equal to 0.808, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.788, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.092, but a value of 0.070 or less means acceptable model fit (Steigler, 2007). CFA results for the SW-PBS/SWPBIS data rendered a significant Chi Square test of base model fit value of 24063.670 with 378 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value relative to the degrees of freedom are considered acceptable for good model fit (Hooper, et al., 2008).

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA of the School-Wide Positive Behavior Supports and Schoolwide Positive Behavioral Interventions and Supports 28-item data can be found in Appendix D Tables 7, 8, and 9. The standardized parameter estimates, or item-factor loading coefficients, for Factor #1 revealed a range from 0.374 to 0.813. The items with the coefficient values less than 0.600 were C3, C5, C8, and C1; but items C7, C2, C9, C4, and C8 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.522 to 0.844. Items E1, E6, E2, E7, and E8 showed standardized parameter estimates of 0.650 or lower; and items E13, E12 E11, E4, and E3 had item-factor loadings larger than 0.650. Factor #3 showed item-factor loadings ranging from 0.571 to 0.891, and while items P3 and P4 had coefficient values larger than 0.800, items P2 and P1 had coefficient values less than 0.800. For Factor #4 the standardized parameter estimates ranged from 0.429 to 0.941. Items F3, F2, and F4 showed item-factor loading values greater than 0.800; but items F1 and F5 showed item-loading values of 0.500 or less.

Factor score coefficients for the SW-PBS/SWPBIS 28-item CFA in Mplus ranged from 0.029 to 0.105 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 8). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.032 to 0.075 on Factor #2, and the Professional Development items revealed values ranging from 0.116 to 0.193 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.018 to 0.157 for the theorized Family Involvement items.

<u>Response-to-Intervention with 28 Items-</u> Confirmatory Factor Analysis was conducted in Mplus on the 2014 Response-to-Intervention data from the Western state; the dataset consisted of the common 28 SIS items, and the 4-factor structure model outlined for the School Implementation Scale was specified. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, et al., 2006). The Comparative Fit Index (CFI) was equal to 0.760, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.737, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.107, but a value of 0.070 or less means acceptable model fit (Steigler, 2007). The CFA results for the RTI data rendered a significant Chi Square test of base model fit value of 4805.496 with 378 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value

relative to the degrees of freedom are considered acceptable for good model fit (Hooper, et al., 2008).

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA of the Response-to-Intervention 28-item data can be found in Appendix D Tables 10, 11, and 12. The standardized parameter estimates, or item-factor loading coefficients, for Factor #1 revealed a range from 0.487 to 0.849. The items with the coefficient values less than 0.700 were C3, C1, C5, and C2; but items C9, C7, C8, C6, and C4 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.577 to 0.906. Items E6, E8, E13, E7, E11, and E2 showed standardized parameter estimates lower than 0.800; and items E4, E3 E1, and E12 had item-factor loadings larger than 0.800. Factor #3 showed item-factor loadings ranging from 0.777 to 1.031, and while items P1 and P3 had coefficient values larger than 1.000, items P2 and P4 had coefficient values between 0.700 and 0.800. For Factor #4 the standardized parameter estimates ranged from 0.296 to 1.136. Items F3, F4, and F2 showed item-factor loading values greater than 1.000; but items F1 and F5 showed item-loading values of 0.600 or less.

Factor score coefficients for the RTI 28-item CFA in Mplus ranged from 0.026 to 0.116 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 11). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.047 to 0.087 on Factor #2, and the Professional Development items revealed values ranging from 0.091 to 0.195 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.012 to 0.174 for the theorized Family Involvement items.

<u>Both States with 28 Items-</u> Confirmatory Factor Analysis was conducted in Mplus on the 2014 data from both the Midwestern and Western states combined, the dataset consisted of the common 28

SIS items, and the 4-factor structure model outlined for the SIS was specified. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, et al., 2006). The Comparative Fit Index (CFI) was equal to 0.822, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.805, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.087, but a value of 0.070 or less means an acceptable model fit (Steigler, 2007). The CFA results for both states' data rendered a significant Chi Square test of base model fit value of 66331.126 with 378 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value relative to the degrees of freedom are considered acceptable for good model fit (Hooper, et al., 2008).

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA of the 28-item data from both the Midwestern and Western states combined can be found in Appendix D Tables 13, 14, and 15. The standardized parameter estimates for Factor #1 revealed a range from 0.420 to 0.842. The items with the coefficient values less than 0.600 were C3, C8, C5, and C1; but items C2, C6, C7, C9, and C4 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.616 to 0.808. Items E13, E1, E6, E11, and E7 showed standardized parameter estimates lower than 0.680; and items E3, E4, E12, and E2 had item-factor loadings larger than 0.680. Factor #3 showed item-factor loadings ranging from 0.659 to 0.910, and while items P3, P1, and P4 had coefficient values larger than 0.800, item P2 had a coefficient value less than 0.700. For Factor #4 the standardized parameter estimates ranged from 0.489 to 1.027. Items F3, F2, and F4 showed item-factor loading values greater than 0.800; but items F1 and F5 showed item-loading values of 0.600 or less.

Factor score coefficients for the 28-item both states data CFA in Mplus ranged from 0.037 to 0.100 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 14). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.028 to 0.091 on Factor #2, and the Professional Development items revealed values ranging from 0.119 to 0.190 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.019 to 0.168 for the theorized Family Involvement items.

The statistical analysis outlined in the research questions must stop there, as the 4-factor structure specified by the School Implementation Scale failed to adequately fit the data. Model fit for the 28-item SIS data from the combined Midwestern and Western state data was required to continue with the remaining research questions. The measurement of invariance from the 4-factor model for the different overall state educational initiatives and across the different multi-tiered systems of support programs of the two states, research questions 2 and 3, would produce invalid results given there is not a strong base model from which to compare. Because measurement invariance of the SIS did not hold to any extent, research questions 4 and 5 cannot be conducted. Neither the reliability of results across the two states or the two different RTI programs in terms of respondent's role, years of experience, or involvement with the implementation team; nor the differences between and patterns within respondents' mean SIS and mean essential element domain scores in terms of respondent demographic characteristics can be validly investigated due to the adequately fit of the theorized 4-factor School Implementation Scale model to the data collected over two states and across two different integrated academic and behavior RTI systems of support.

Exploratory Factor Analysis Specifying 4 Factors in SPSS

Midwestern State Collaborative Work with 28 items- Exploratory Factor Analysis specifying that 4 factors be extracted was conducted in SPSS on the Midwestern state's Collabortive Work dataset consisting of the 28 common School Implementation Scale items employing PAF and Promax rotation but specifying 4 factors as the theorized factor structure of the SIS outlined. The theorized 4-factor structure accounted for a cumulative 57% of the variance. Table 9 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 10.9 accounting for 39% of variance; a smaller yet prominent factor with an Eigenvalue of 2.1 accounting for 7% of the variance; and 2 much smaller factors with Eigenvalues of 1.7 (6%) and 1.2 (4%).

Table 9. Midwestern CW 28 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.925	39.017	39.017
2	2.062	7.365	46.382
3	1.677	5.988	52.369
4	1.168	4.171	56.541

Table 10. Midwestern CW 28 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	10	#2	2	2	2			4		0.865
		#4	2						2 #2	
		#1			1				1 #2	
		#3			1				1 #2	
School Culture	9	#1	2	3	2	1		1		0.851
		#2		1					1 #1	
Family Involvement	5	#3	3					2		0.832
		#1		1					1 #3	
		#4				1			1 #3	
Professional Development	4	#2		2	1			3	3 #4	0.823
		#1		1					1 #4	
Total SIS	28									0.939

Table 10 displays a summary of the rotated item-factor structure for the Midwestern state's Collaborative Work 28-item version of the SIS instrument resulting from the SPSS EFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 1. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 6 items loaded on Factor #2 with 2 excellent, 2 very good, and 2 good item-factor loadings. Two of the theorized Evidence-Based Practices items loaded on Factor #4 with excellent item-loading values, but showed fair cross-loading values on Factor #2. One theorized Evidence-Based Practices item had a good item-factor loading on Factor #1 with a fair cross-loading value on Factor #2, and the final theorized Evidence-Based Practices item had a good item-factor loading on Factor #3 with a fair cross-loading value on Factor #2. Cronbach's Alpha for the Evidence-Based Practices factor was 0.865, and the elimination of any of the 10 items did not increase the internal consistency measure for the factor.

Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #1 with 2 excellent, 3 very good, 2 good, and 1 fair item-factor loadings; and the final School Culture item showed a very good item-loading value on Factor #2 with a poor cross-loading value on Factor #1. Cronbach's Alpha for the School Culture factor was 0.851, and the elimination of item C3 (I have a clear understanding of the State Standards for my grade/subject) would increase the internal consistency measure for the factor to 0.853.

Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #3. One Family Involvement theorized item had a very good loading on Factor #1 with a fair cross-loading on Factor #3, and the final Family Involvement item had a fair item-factor loading value on Factor #4 with a fair cross-loading on Factor #3. Cronbach's Alpha for the Family Involvement factor was 0.832. However, the elimination of item F1 (I consider my students' backgrounds when planning instruction) would increase the internal consistency measure for the factor to 0.835, and the elimination of item F5 (I think my school does a good job of including parents as team members in data-based decision making) would increase Cronbach's alpha for the Family Involvement factor to 0.844.

Of the 4 items theorized to be included in the Professional Development factor, 3 loaded on Factor #2 with 2 very good and 1 good item-factor loading values; and the final theorized Professional Development item had a very good item-factor loading on Factor #1. The 4 Professional Development items experienced poor to fair item-factor loading values on Factor #4. Cronbach's Alpha for the Professional Development factor was 0.823, and the elimination of any of the 4 items did not increase the internal consistency measure for the factor.

The Cronbach's Alpha for the entire 28-item School Implementation Scale was 0.939, which showed a high level of internal consistency among the 28 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 1. It was revealed that the elimination of any of the 28 items would not increase the internal consistency measure for the scale, but Cronbach's Alpha remained at 0.939 if item C5 (I have the technology and resources that I need to provide effective instruction) was deleted. The alpha decreased slightly to 0.938 if items C3, C8, or E1 were removed; and decreased to 0.937 if items C1, C6, E2, E4, E7, E8, E12, E13, P2, F1, F2, F3, or F5 were deleted.

The rotated item-factor structure for the SPSS CFA on the Midwestern CW 28-item dataset seemed to suggest that Factor #1 was School Culture, Factor #2 was Evidence-Based Practices, Factor #3 was Family Involvement, and Factor #4 was Professional Development. The Factor Correlation Matrix presented in Table 11 showed that the highest correlated factors were School Culture and Evidence-Based Practices at 0.664, followed by Evidence-Based Practices and Professional Development at 0.634, and Family Involvement and Professional Development at 0.619. School Culture and Professional Development were correlated at 0.603, Family Involvement and Evidence-Based Practices at 0.589, and the lowest correlation between factors was School Culture and Family Involvement at 0.563.

Table 11. Midwestern CW 28 Items- EFA specifying 4 factors Factor Correlation Matrix

	School Culture	Evidence-Based Practices	Family Involvement	Professional Development
School Culture	1.000	.664	.563	.603
Evidence-Based Practices	.664	1.000	.589	.634
Family Involvement	.563	.589	1.000	.619
Professional Development	.603	.634	.619	1.000

<u>Western State Effective Behavioral and Instructional Support Systems with 28 Items-</u>

Exploratory Factor Analysis specifying the extraction of 4 factors was conducted in SPSS on the Western state's EBISS dataset consisting of the 28 common School Implementation Scale items, and analysis employed PAF and Promax rotation but specified 4 factors as the theorized factor structure of the SIS outlined. The theorized 4-factor structure accounted for a cumulative 57% of the variance. Table 12 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 10.7 and accounting for 38% of variance; a smaller yet prominent factor with an Eigenvalue of 2.8 and accounting for almost 9% of the variance; and 2 much smaller factors with Eigenvalues of 1.4 (5%) and 1.2 (4%).

Table 12. Western EBISS 28 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.715	38.267	38.267
2	2.767	9.881	48.148
3	1.374	4.908	53.057
4	1.170	4.179	57.236

Table 13 displays the rotated item-factor structure for the Western state's EBISS common 28 items of the SIS instrument resulting from the SPSS EFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 2. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 6 items loaded on Factor #3 with 3 excellent, 2 very good, and 1 good item-factor loadings. Three of the theorized Evidence-Based Practices items loaded on Factor #2 with excellent or good item-loading values, but showed fair and good cross-loading values on Factor #3; and the final theorized Evidence-Based Practices item had a very good item-factor loading on Factor #1 with a good cross-loading value on Factor #3. Cronbach's Alpha for the Evidence-Based Practices factor was 0.893, and the elimination of any of the 10 items did not increase the internal consistency measure for the factor.

Table 13. Western EBISS 28 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	10	#3	3	2	1				4	0.893
		#2	2		1			3 #3		
		#1		1				1 #3		
School Culture	9	#1	3	1	1	2	1		1	0.818
		#2		1				1 #1		
Family Involvement	5	#4	3						2	0.802
		#1			1			1 #4		
		#3			1			1 #4		
Professional Development	4	#2	1		1				2	0.806
		#3		1	1			2 #2		
Total SIS	28									0.937

Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #1 with 3 excellent, 1very good, 1 good, 2 fair, and 1 poor item-factor loadings; and the final School Culture item showed a very good item-loading value on Factor #2 with a poor cross-loading value on Factor #1. Cronbach's Alpha for the School Culture factor was 0.818, and the elimination of item C3 (I have a clear understanding of the Common Core State Standards) or item C5 (I have the technology and resources that I need to provide effective instruction) would increase the measure of internal consistency to 0.819.

Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #4. One theorized Family Involvement item had a good loading on Factor #1, but had a cross-loading on Factor #4 that was non-interpretable at less than 0.320 (Comrey & Lee, 1992). The final Family Involvement item had a good item-factor loading value on Factor #3 with a fair cross-loading on Factor #4. Cronbach's Alpha for the Family Involvement factor was 0.802, and the elimination of item F5 (I think that my school does a good job of including parents as team members in data-based decision making) would increase the internal consistency measure for the factor to 0.867.

Of the 4 items theorized to be included in the Professional Development SIS factor, 2 loaded on Factor #2 with 1 excellent and 1 good item-factor loading values. The other 2 theorized Professional Development items showed very good and good item-factor loadings on Factor #3 with good and fair cross-loading values on Factor #2. Cronbach's Alpha for the Professional Development factor was 0.806, and the elimination of any of the 4 items did not increase the internal consistency measure for the factor.

The Cronbach's Alpha for the entire 28-item School Implementation Scale was 0.937, which showed a high level of internal consistency among the 28 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 2. It was revealed that the elimination of item C5 (I have the technology and resources that I need to provide effective instruction) would increase the internal consistency measure for the scale to 0.938. The removal of items C1 (I can summarize my school's shared vision/mission) or F5 (I think that my school does a good job of including parents as team members in data-based decision making) would not change the Cronbach's Alpha value from 0.937.

The rotated item-factor structure for the SPSS EFA specifying 4 factors on the Western EBISS 28 item dataset seemed to suggest that Factor #1 was School Culture, Factor #2 was Professional Development, Factor #3 was Evidence-Based Practices, and Factor #4 was Family Involvement. The Factor Correlation Matrix in Table 14 shows that the highest correlated factors were Evidence-Based Practices and Professional Development at 0.676, followed by Evidence-Based Practices and School Culture at 0.636, and Family Involvement and Evidence-Based Practices at 0.628. Family Involvement and Professional Development were correlated at 0.602, Professional Development and School Culture at 0.426, and the lowest correlation between factors was School Culture and Family Involvement at 0.375.

Table 14. Western EBISS 28 Items- EFA specifying 4 factors Factor Correlation Matrix

	School Culture	Professional Development	Evidence-Based Practices	Family Involvement
School Culture	1.000	.426	.636	.375
Professional Development	.426	1.000	.676	.602
Evidence-Based Practices	.636	.676	1.000	.628
Family Involvement	.375	.602	.628	1.000

<u>School-Wide Positive Behavior Support & Schoolwide Positive Behavior Intervention Supports with 28 Items-</u> Exploratory Factor Analysis specifying the extraction of 4 factors was conducted in SPSS on the tiered systems of support program of School-Wide Positive Behavior Support from the Midwestern state and Schoolwide Positive Behavior Interventions and Supports from the Western state. The SW-PBS/SWPBIS dataset from both states combined consisted of the 28 common School Implementation Scale items, and the CFA employed PAF and Promax rotation but specified 4 factors as the theorized factor structure of the SIS outlined. The theorized 4-factor

structure accounted for a cumulative 57% of the variance. Table 15 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 11.2 accounting for 40% of variance; a smaller yet prominent factor with an Eigenvalue of 2.3 accounting for 8% of the variance; and 2 much smaller factors with Eigenvalues of 1.4 (5%) and 1.1 (almost 4%).

Table 15. SW-PBS/SWPBIS 28 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	11.175	39.912	39.912
2	2.262	8.077	47.989
3	1.384	4.941	52.930
4	1.078	3.851	56.781

Table 16. SW-PBS/SWPBIS 28 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	10	#3	2	3	2				4	0.893
		#1	3	1				4 #3		
School Culture	9	#2	2		2	1			4	0.837
		#3	1		2			3 #2		
		#1			1					
Family Involvement	5	#4	3						2	0.814
		#1		1				1 #4		
		#2		1				1 #4		
Professional Development	4	#1		2	1				1	0.805
		#4		1				1 #1		
Total SIS	28									0.942

Table 16 displays a summary of the rotated item-factor structure of the 28 common SIS items for the School-Wide Positive Behavior Supports from the Midwestern state and the Schoolwide Positive Behavior Interventions and Supports from the Western state dataset resulting from the SPSS EFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 3. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 6 items loaded on Factor #3 with 2 excellent, 3 very good, and 2 good item-factor loadings. Four of the theorized Evidence-Based Practices items loaded on Factor #1 with excellent and very good item-loading values, but these items showed poor, fair, and good cross-loading values on Factor #3. Cronbach's Alpha for the Evidence-Based Practices factor was 0.893, and the elimination of any of the 10 items did not increase the internal consistency measure for the factor.

Five of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #2 with 2 excellent, 2 good, and 1 fair item-factor loadings. Three theorized School Culture items showed good and very good item-loading values on Factor #3 with fair and good cross-loading values on Factor #2, and the final theorized School Culture item showed a good loading on Factor #1 and a non-interpretable (less than 0.320) cross-loading value on Factor #2 (Comrey & Lee, 1992). Cronbach's Alpha for the School Culture factor was 0.837, and the elimination of any of the 9 items did not increase the internal consistency measure for the factor. However, the removal of item C3 (I have a clear understanding of the State Standards for my grade/subject) would not change the 0.837 value of the internal consistency measure.

Of the 5 items theorized to be included in the Family Involvement SIS domain items, 3 items showed excellent item-factor loadings on Factor #4. One theorized Family Involvement item had a very good loading on Factor #1 with a poor cross-loading on Factor #4, and the final Family Involvement item had a very good item-factor loading value on Factor #2 with a poor cross-loading value on Factor #4. Cronbach's Alpha for the Family Involvement factor was 0.814, and the elimination of item F5 (I think my school does a good job of including parents as team members in data-based decision making) would increase the internal consistency measure for the factor to 0.848.

Of the 4 items theorized to be included in the Professional Development SIS domain, 3 loaded on Factor #1 with 2 very good and 1 good item-factor loading values; and the final theorized Professional Development item had a very good item-factor loading on Factor #4 with a good cross-loading value on Factor #1. Cronbach's Alpha for the Professional Development factor was 0.805, and the elimination of any of the 4 items did not increase the internal consistency measure for the factor.

The Cronbach's Alpha for the entire 28-item School Implementation Scale was 0.942, which showed a high level of internal consistency among the 28 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 3. It was revealed that the elimination of item C5 (I have the technology and resources that I need to provide effective instruction) would increase the internal consistency measure to 0.943, but the deletion of items C8, C3, C6, C1, C4, or F5 would only slightly decrease Cronbach's Alpha to 0.941.

The rotated item-factor structure for the SPSS EFA specifying 4 factors on the SW-PBS/SWPBIS 28-item dataset seemed to suggest that Factor #1 was Professional Development, Factor #2 was School Culture, Factor #3 was Effective Practices, and Factor #4 was Family Involvement. The Factor Correlation Matrix in Table 17 shows that the highest correlated factors seem to be Professional Development and Effective Practices at 0.693, followed by Effective Practices and School Culture at 0.680, and Family Involvement and Professional Development at 0.627. Effective Practices and Family Involvement were correlated at 0.614, School Culture and

Table 17. SW-PBS/SWPBIS 28 Items- EFA specifying 4 factors Factor Correlation Matrix

	Professional Development	School Culture	Evidence-Based Practices	Family Involvement
Professional Development	1.000	0.501	0.693	0.627
School Culture	0.501	1.000	0.680	0.360
Evidence-Based Practices	0.693	0.680	1.000	0.614
Family Involvement	0.627	0.360	0.614	1.000

Western State Response-to-Intervention with 28 Items- Exploratory Factor Analysis specifying the extraction of 4 factors was conducted in SPSS on the tiered systems of support program of Response-to-Intervention dataset from the Western state. The Western RTI dataset consisted of the 28 common School Implementation Scale items, and the EFA employed PAF and Promax rotation but specified 4 factors as the theorized factor structure of the SIS items outlined. The theorized 4-factor structure accounted for a cumulative 59% of the variance. Table 18 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 11.0 accounting for 40% of variance; one smaller yet prominent factor with an Eigenvalue of 2.8 accounting for 10% of the variance; and 2 much smaller factors with Eigenvalues of 1.5 (5%) and 1.2 (4%).

Table 18. RTI 28 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.982	39.221	39.221
2	2.800	10.000	49.221
3	1.461	5.219	54.440
4	1.232	4.399	58.839

Table 19 displays the rotated item-factor structure of the common 28 SIS items for the Western state's RTI tiered system of support program resulting from the SPSS CFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 4. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 4 items loaded on Factor #2 with 2 excellent and 2 very good item-factor loadings. Four of the theorized Evidence-Based Practices items loaded on Factor #1 with excellent, very good, or good item-loading values but showed fair and very good cross-loading values on Factor #2. One theorized Evidence-Based Practices item had an excellent item-factor loading on Factor #3 with a poor cross-loading value on Factor #2, and the final theorized Evidence-Based Practices item had an excellent item-factor loading on Factor #4 with a fair cross-loading value on Factor #2. Cronbach's Alpha for the Evidence-Based Practices factor was 0.889, and the elimination of none of the 10 items would increase the internal consistency measure for the factor.

Table 19. RTI 28 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	10	#2	2	2					6	0.889
		#1	1	2	1			4 #2		
		#3	1					1 #2		
		#4	1					1 #2		
School Culture	9	#3	3	1	1				5	0.828
		#4		1		1	1	1 #3		
		#2				1		1 #3		
Family Involvement	5	#1	3	1					1	0.797
		#4			1					
Professional Development	4	#4		1	1				2	0.829
		#2	1	1				2 #4		
Total SIS	28									0.940

Five of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #3 with 3 excellent, 1very good, and 1 good item-factor loadings. Three theorized School Culture items showed very good, fair, or poor item-factor loading values on Factor #4 with poor or non-interpretable (less than 0.320) cross-loading values on Factor #3. The final theorized School Culture item had a very good item-factor loading on Factor #2 and showed a fair cross-loading value on Factor #3. Cronbach's Alpha for the School Culture factor was 0.828, and the elimination of item C5 (I have the technology and resources that I need to provide effective instruction) would increase the measure of internal consistency for the factor to 0.829.

Of the 5 items theorized to be included in the Family Involvement SIS factor, 4 items showed excellent or very good item-factor loadings on Factor #1. The final theorized Family Involvement item had a good loading on Factor #4 with a cross-loading on Factor #1 that was non-interpretable (less than 0.320). Cronbach's Alpha for the Family Involvement factor was 0.797, and the elimination of item F5 (I think my school does a good job of including parents as team members in data-based decision making) would increase the measure of internal consistency to 0.853.

Of the 4 items theorized to be included in the Professional Development SIS factor, 2 loaded on Factor #4 with 1 very good and 1 good item-factor loading values. The other 2 theorized Professional Development items showed excellent and very good item-factor loadings on Factor #2 with good cross-loading values on Factor #4. Cronbach's Alpha for the Professional Development factor was 0.829, and the elimination of any of the 4 items would not increase the measure of internal consistency for the factor.

The Cronbach's Alpha for the entire 28-item School Implementation Scale was 0.940, which showed a high level of internal consistency among the 28 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 4. It was revealed that the removal of item C5 (I have the technology and resources that I need to provide effective instruction) would increase Cronbach's Alpha for the factor to 0.942. The elimination of items C1 (I can summarize my schools' shared vision/mission), C4 (I think my school has an effective process in place to identify available resources), or F5 (I think my school does a good job of including parents as team members in data-based decision making) would not change the measure of internal consistency from 0.940. Cronbach's Alpha would decrease only slightly to 0.939 if items C3, C6, C2, E6, F2, or C9 were deleted from the scale.

The rotated item-factor structure for the SPSS EFA specifying 4 factors of the Western RTI 28-item dataset seemed to suggest that Factor #1 was Family Involvement, Factor #2 was Evidence-Based Practices, Factor #3 was School Culture, and Factor #4 was Professional Development. The Factor Correlation Matrix presented in Table 20 shows that the highest correlated factors were Evidence-Based Practices and Family Involvement at 0.685, followed by Professional Development and School Culture at 0.586, and Family Involvement and Professional Development at 0.581. Evidence-Based Practices and Professional Development were correlated at 0.540, Evidence-Based Practices and School Culture at 0.411, and the lowest correlation between factors was School Culture and Family Involvement at 0.403.

Table 20. RTI 28 Items- EFA specifying 4 factors Factor Correlation Matrix

	Family Involvement	Evidence-Based Practices	School Culture	Professional Development
Family Involvement	1.000	0.685	0.403	0.581
Evidence-Based Practices	0.685	1.000	0.411	0.540
School Culture	0.403	0.411	1.000	0.586
Professional Development	0.581	0.540	0.586	1.000

Both States with 28 items- Exploratory Factor Analysis specifying 4 factors be extracted and employing PAF and Promax rotation was conducted in SPSS on the combined Midwestern and Western states' data. The dataset consisted of the 28 common items of the original 31 School Implementation Scale items, and 4 factors were specified as the theorized factor structure of the SIS outlined. The theorized 4-factor structure of the SIS accounted for a cumulative 58% of the variance. Table 21 shows the Eigenvalues and percent of variance accounted for by each of the 4 factors- one very large factor with an Eigenvalue of 11.5 accounting for 41% of the variance, a smaller yet prominent factor with an Eigenvalue of 2.2 accounting for almost 8% of the variance, and 2 much smaller factors with Eigenvalues of 1.4 (almost 5%) and 1.1 (almost 4%).

Table 21. Both States 28 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	11.534	41.192	41.192
2	2.185	7.803	48.996
3	1.355	4.839	53.834
4	1.082	3.864	57.699

Table 22. Both States 28 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	10	#1	4	2	2				2	0.887
		#4		1		1		2 #1		
School Culture	9	#2	3	2	3	1			1	0.851
		#1			1			1 #2		
Family Involvement	5	#3	3						2	0.831
		#1			1			1 #3		
		#2		1				1 #3		
Professional Development	4	#4							4	0.829
		#1		3				3 #4		
		#2			1			1 #4		
Total SIS	28									0.945

Table 22 displays the rotated item-factor structure for the combined Midwestern and Western state data consisting of the 28 common School Implementation Scale items which resulted from the SPSS EFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 5. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 8 items loaded on Factor #1 with 4 excellent, 2 very good, and 2 good item-factor loadings. Two of the theorized Evidence-Based Practices items loaded on Factor #4 with very good and fair item-loading values, but showed fair cross-loading values on Factor #1. The Cronbach's Alpha for the Evidence-Based Practices factor was 0.887, and elimination of none of the 28 items would increase the measure of internal consistency. However, the removal of item E4 (I review universal screening data at least three times a year for every student that I support) would not change Cronbach's Alpha for the Evidence-Based Practices factor.

Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #2 with 3 excellent, 2 very good, 3 good, and 1 fair item-factor loadings. The final theorized School Culture item showed a good item-factor loading value on Factor #1 with a poor cross-loading value on Factor #2. Cronbach's Alpha for the School Culture factor was 0.851, and the elimination of none of the 9 items would increase the measure of internal consistency. The removal of item C3 (I have a clear understanding of the State Standards for my grade/subject) would only decrease Cronbach's Alpha slightly to 0.850.

Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #3. One theorized Family Involvement item had a good loading on Factor #1 with a fair cross-loading on Factor #3, and the final theorized Family Involvement item had a very good item-factor loading on Factor #2 with a poor cross-loading on Factor #3. Cronbach's Alpha for the Family Involvement factor was 0.831, but the removal of item F5 (I think my school does a good job of including parents as team members in data-based decision making) would increase the measure of internal consistency to 0.857. The elimination of item F1 (I consider my students' backgrounds when planning instruction) would only decrease Cronbach's Alpha slightly to 0.830.

Of the 4 items theorized to be included in the Professional Development SIS factor, 3 items showed very good item-factor loadings on Factor #1 with poor, fair, and good cross-loadings on Factor #4. The final theorized Professional Development item showed a good item-factor loading on Factor #2 with a poor cross-loading value on Factor #4. Cronbach's Alpha for the Professional Development factor was 0.829, and the elimination of any of the 4 Professional Development items did not increase the measure of internal consistency.

The Cronbach's Alpha for the entire 28-item School Implementation Scale was 0.945, which showed a high level of internal consistency among the 28 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 5. It was revealed that the elimination of none of the 28 items would increase the alpha, but deleting item C5 (I have the technology and resources that I need to provide effective instruction) would not cause a change in the internal consistency. The removal of items C8 (I feel that my administrators are committed to implementing evidence-based instructional practices), C3 (I have a clear understanding of the State Standards for my grade/subject), C6 (I have the time necessary to analyze students data and problem-solve with my colleagues), C1 (I can summarize my schools' shared vision/mission), or E4 (I review universal screening data at least three times a year for every student that I support) would decrease the Cronbach's Alpha slightly to 0.944. The removal of items E2, P2, P4, E8, E7, E1, E13, E11, C2, E6, C4, C7, F2, F3, F4, F5, or F1 would decrease the measure of internal consistency to 0.943.

The rotated item-factor structure for the SPSS EFA specifying 4 factors of both states' 28-item datasets combined seemed to suggest that Factor #1 was Effective Practices, Factor #2 was School Culture, Factor #3 was Family Involvement, and Factor #4 was Professional

Development. The Factor Correlation Matrix presented in Table 23 shows that the highest correlated factors were Effective Practices and Family Involvement at 0.670, followed by Effective Practices and School Culture at 0.641, and Family Involvement and School Culture at 0.540. School Culture and Professional Development were correlated at 0.482, Effective Practices and Professional Development at 0.451, and the lowest correlation between factors was Professional Development and Family Involvement at 0.344.

Table 23. Both States 28 Items- EFA specifying 4 factors Factor Correlation Matrix

	Evidence-Based Practices	School Culture	Family Involvement	Professional Development
Evidence-Based Practices	1.000	0.641	0.670	0.451
School Culture	0.641	1.000	0.540	0.482
Family Involvement	0.670	0.540	1.000	0.344
Professional Development	0.451	0.482	0.344	1.000

Exploratory Factor Analysis without Specifying Number of Factors to Extract in SPSS

Midwestern State Collaborative Work with 28 Items- Exploratory Factor Analysis (EFA) without specifying the number of factors to extract was conducted in SPSS on the 2013-2014 Midwestern state's Collaborative Work dataset consisting of only 28 of the original 31 School Implementation Scale instrument items which are common across the datasets of both states. The 28 SIS items included 10 items in the domain of Effective Practices, 9 items under the domain of School Culture, 5 items in the Family Involvement domain, and 4 items under the Professional Development domain as the theorized 4-factor structure of the School Implementation Scale items outlined. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.945, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and the Bartlett's Test of Sphericity rendered an Approximate Chi

Square value of 46928.544 (378 degrees of freedom and p<0.001). Six factors were extracted with Eigenvalues larger than 1.0 and accounted for a cumulative 64% of the variance. Table 24 shows the Eigenvalues and percent of variance for each of the 6 extracted factors- one very large factor with an Eigenvalue of 10.9 accounting for 39% of variance; a smaller yet prominent factor with an Eigenvalue of 2.1 accounting for 7% of the variance; an even smaller factor with an Eigenvalue of 1.7 (6%); and 3 much smaller factors with Eigenvalues of 1.2 (4%), 1.1 (almost 4%), and 1.0 (almost 4%).

Table 24. Midwestern CW 28 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.925	39.017	39.017
2	2.062	7.365	46.382
3	1.677	5.988	52.369
4	1.168	4.171	56.541
5	1.076	3.843	60.384
6	1.049	3.747	64.131

Table 25 displays a summary of the rotated item-factor structure for the Midwestern state's Collaborative Work 28-item version of the School Implementation Scale resulting from the SPSS EFA, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 6. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 4 items loaded on Factor #3 with 2 excellent, 1 very good, and 1 good item-factor loadings. Four of the theorized Evidence-Based Practices items loaded on Factor #5 with 2 excellent, 1 very good, and 1 good item-loading values. One theorized Evidence-Based Practices item showed a good on Factor #1, and the final theorized Evidence-Based Practices item had a good loading on Factor #6. Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #1 with 1 excellent, 3 very good, and 4 good item-factor loadings; and the final theorized School Culture item had a good item-loading value on Factor #3. Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #2. One theorized Family Involvement item had a very good loading on Factor #1, and the final Family Involvement item had a good loading on Factor #5. Of the 4 items theorized to be included in the Professional Development SIS factor, all loaded on Factor #4 with 3 excellent and 1 very good item-factor loadings.

Table 25. Midwestern CW 28 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#3	2	1	1		
		#5	2	1	1		
		#1			1		
		#6			1		
School Culture	9	#1	1	3	4		
		#3			1		
Family Involvement	5	#2	3				
		#1		1			
		#5			1		
Professional Development	4	#4	3	1			

<u>Western State Effective Behavioral and Instructional Support Systems with 28 Items-</u>

Exploratory Factor Analysis without specifying the number of factors to extract was conducted in SPSS on the 2013-2014 Western state's EBISS dataset consisting of 28 of the original 31 School Implementation Scale instrument items which were common across the datasets of both states. The 28 SIS items consisted of 10 items in the domain of Effective Practices, 9 items under the domain of School Culture, 5 items in the Family Involvement domain, and 4 items under the Professional Development domain as the theorized item-factor structure of the School Implementation Scale outlined. The examination of factorability showed a Kaiser-Meyer-Olkin

Measure of Sampling Adequacy equal to 0.939, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and the Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 21201.408 (378 degrees of freedom and p<0.001). Five factors were extracted with Eigenvalues larger than 1.0 and accounted for a cumulative 61% of the variance. Table 26 shows the Eigenvalues and percent of variance for each of the 5 extracted factors- one very large factor with an Eigenvalue of 10.7 accounting for 38% of variance; a smaller yet prominent factor with an Eigenvalue of 2.8 accounting for almost 10% of the variance; an even smaller factor with an Eigenvalue of 1.4 (5%); and 2 much smaller factors with Eigenvalues of 1.2 (4%) and 1.0 (almost 4%).

Table 26. Western EBISS 28 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.715	38.267	38.267
2	2.767	9.881	48.148
3	1.374	4.908	53.057
4	1.170	4.179	57.236
5	1.037	3.704	60.940

Table 27 displays the rotated item-factor structure resulting from the EFA for the Western state's EBISS 28-item version of the School Implementation Scale, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 7. Of the 10 items theorized to be in the factor of Evidence-Based Practices, 4 items loaded on Factor #4 with 2 excellent, 1 very good, and 1 good item-factor loadings (Comrey & Lee, 1992). Three of the theorized Evidence-Based Practices items loaded on Factor #5 with 2 excellent and 1 very good item-loading values. Two theorized Evidence-Based Practices items had excellent loadings on Factor #2, and the final theorized Evidence-Based Practices item had a very good loading on Factor #1.

Table 27. Western EBISS 28 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#4	2	1	1		
		#5	2	1			
		#2	2				
		#1		1			
School Culture	9	#1	3	1	1	1	1
		#2		1			
		#5			1		
Family Involvement	5	#3	3				
		#1			1		
		#4			1		
Professional Development	4	#2	1	1			
		#1			1		
		#5		1			

Seven of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #1 with 3 excellent, 1 very good, 1 good, 1 fair, and 1 poor item-factor loadings. One of the theorized School Culture items showed a very good loading value on Factor #2, and the final theorized School Culture item showed a good loading on Factor #5. Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #3. One Family Involvement theorized item had a good loading on Factor #1, and the final Family Involvement item had a good loading on Factor #4. Of the 4 items theorized to be included in the Professional Development SIS factor, 2 items loaded on Factor #2 with 1 excellent and 1 very good item-loading values. One of the theorized Professional Development items showed a good loading on Factor #1, and the final Professional Development item showed a very good loading on Factor #5.

School-Wide Positive Behavior Support & Schoolwide Positive Behavior Intervention Supports with 28 Items- Exploratory Factor Analysis without specifying the number of factors to extract was conducted in SPSS on the tiered systems of support program of School-Wide Positive Behavior Support from the Midwestern state and Schoolwide Positive Behavior Interventions and Support from the Western state. The SW-PBS/SWPBIS dataset from both states combined consisted of the 28 common School Implementation Scale items, and the EFA employed PAF and Promax rotation. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.946, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial between items were small, and the Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 23762.346 (378 degrees of freedom and $p<0.001$). Four factors were extracted and accounted for a cumulative 57% of the variance. Table 28 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 12.9 accounting for 40% of variance; a smaller yet prominent factor with an Eigenvalue of 2.5 accounting for 8% of the variance; and 2 much smaller factors with Eigenvalues of 1.6 (5%) and 1.4 (4%).

Table 28. SW-PBS/SWPBIS 28 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	12.858	40.186	40.186
2	2.510	7.843	48.029
3	1.591	4.973	53.001
4	1.382	4.320	57.322

Table 29 displays a summary of the rotated item-factor structure for the School-Wide Positive Behavior Supports from the Midwestern state and the Schoolwide Positive Behavior Interventions and Supports from the Western state tiered systems of support programs data for the 28 common items of the SIS instrument resulting from the SPSS EFA. The full Rotated Factor Structure Matrix can be found in Appendix E Table 8. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 6 items loaded on Factor #3 with 3 excellent, 2 very good, and 1 good item-factor loadings (Comrey & Lee, 1992).

Table 29. SW-PBS/SWPBIS 28 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#3	3	2	1		
		#1	2	2			
School Culture	9	#2	3		1	2	
		#3	1	1			
		#1				1	
Family Involvement	5	#4	3				
		#1			1		
		#2	1				
Professional Development	4	#1	2		1		
		#4	1				

Four of the theorized Evidence-Based Practices items loaded on Factor #1 with 2 excellent and 2 very good item-loading values. Six of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #2 with 3 excellent, 1 good, and 2 fair item-factor loadings. Two theorized School Culture items showed excellent and very good loadings on Factor #3, and the final School Culture item showed a fair loading on Factor #1. Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #4. One Family Involvement theorized item had a good loading on Factor #1, and the final Family Involvement item had an excellent loading on Factor #2. Of the 4 items theorized to be included in the Professional Development SIS factor, 3 loaded on Factor #1 with 2 excellent and 1 good item-factor loading values; and the final theorized Professional Development item showed an excellent item-factor loading on Factor #4.

Response-to-Intervention with 28 Items- Exploratory Factor Analysis without specifying the number of factors to extract was conducted in SPSS on the tiered system of support program of Response-to-Intervention from the Western state. The Western RTI dataset consisted of the 28 common School Implementation Scale items, and the EFA employed PAF and Promax rotation. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.928, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and the Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 4348.542 (378 degrees of freedom and $p<0.001$). Five factors were extracted with Eigenvalues greater than 1, and accounted for a cumulative 63% of the variance. Table 30 shows the Eigenvalues and percent of variance for each of the 5 extracted factors- one very large factor with an Eigenvalue of 10.9 accounting for 39% of variance; one smaller yet prominent factor with an Eigenvalue of 2.8 accounting for 10% of the variance; an even smaller factor with an Eigenvalue of 1.5 (5%); and 2 much smaller factors with Eigenvalues of 1.2 (4%), and 1.0 (almost 4%).

Table 30. RTI 28 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	10.982	39.221	39.221
2	2.800	10.000	49.221
3	1.461	5.219	54.440
4	1.232	4.399	58.839
5	1.032	3.686	62.525

Table 31 displays the rotated item-factor structure for the Western state's RTI tiered system of support program data from the 28 common items of the SIS instrument resulting from the SPSS EFA, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 9. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 4 items loaded on Factor #1 with 1 excellent, 2 very good, and 1 fair item-factor loadings according to Comrey & Lee (1992). Two of the theorized Evidence-Based Practices items loaded on Factor #4 with excellent or very good item-loading values, 2 items had excellent loadings on Factor #5, 1 item showed an excellent loading on Factor #2, and the final theorized Evidence-Based Practices item showed an excellent loading on Factor #3.

Table 31. RTI 28 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#1	1	2		1	
		#4	1	1			
		#5	2				
		#2	1				
		#3	1				
School Culture	9	#2	3	1	1		
		#3		1		2	
		#4			1		
Family Involvement	5	#1	3				
		#3			1		
		#5	1				
Professional Development	4	#3		1	1		
		#2	1		1		

Four of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #2 with 3 excellent, 1 very good, and 1 good item-factor loadings. Three theorized School Culture items showed very good or fair loading values on Factor #3, and the final School Culture item showed a good loading on Factor #4. Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items had excellent item-factor loadings on Factor #1, 1 item had a good loading on Factor #3, and the final theorized Family Involvement item had an excellent loading on Factor #5. Of the 4 items theorized to be included in the Professional Development SIS factor, 2 loaded on Factor #3 with very good and good item-factor loading values. The other 2 theorized Professional Development items showed excellent and good item-factor loadings on Factor #2.

Both States with 28 Items- Exploratory Factor Analysis using Principle Axis Factoring for the factor extraction method and Promax rotation for oblique or correlated factors was conducted in SPSS without specifying the number of factors to extract on the combined Midwestern state's Collaborative Work and Western state's EBISS data for the common 28 of the original 31 School Implementation Scale items. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was equal to 0.954, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and the Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 69592.948 (378 degrees of freedom and $p<0.001$). Four factors were extracted with Eigenvalues larger than 1.0 and accounted for a cumulative 58% of the variance. Table 32 shows the Eigenvalues and percent of variance accounted for by each of the 4 extracted factors- one very large factor with Eigenvalue of 11.5 accounting for 41% of variance, a smaller yet prominent factor with Eigenvalue of 2.2 accounting for almost 8% of the variance, and 2 much smaller factors with Eigenvalues of 1.4 (almost 5%) and 1.1 (almost 4%).

Table 32. Both States 28 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	11.534	41.192	41.192
2	2.185	7.803	48.996
3	1.355	4.839	53.834
4	1.082	3.864	57.699

Table 33 displays a summary of the rotated item-factor structure resulting from the SPSS EFA for the combined Midwestern and Western state data consisting of the 28 common SIS items, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 10. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 8 items loaded on Factor #1 with 4 excellent, 2 very good, and 2 good item-factor loadings (Comrey & Lee, 1992). Two of the theorized Evidence-Based Practices items loaded on Factor #4 with very good and fair good item-loading values. Eight of the 9 items theorized to be in the factor of School Culture loaded on Factor #2 with 2 excellent, 2 very good, 3 good, and 1 fair item-factor loadings. The final theorized School Culture item showed a good loading value on Factor #1. Of the 5 items theorized to be in the SIS factor of Family Involvement, 3 items had excellent loadings on Factor #3. One theorized Family Involvement item showed a good loading on Factor #1, and the final Family Involvement item showed a very good loading on Factor #2. Of the 4 items theorized to be in the SIS factor of Professional Development, 3 items had very good loadings on Factor #1, and the final theorized Professional Development item had a good loading on Factor #2.

Table 33. Both States 28 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#1	4	2	2		
		#4		1		1	
School Culture	9	#2	2	2	3	1	
		#1			1		
Family Involvement	5	#3	3				
		#1			1		
		#2		1			
Professional Development	4	#3	3				
		#1			1		

Midwestern State Collaborative Work with 37 Items- Exploratory Factor Analysis without specifying the number of factors to extract was conducted in SPSS on the 2013-2014 Midwestern state's complete dataset including all survey items as administered. Thirty-seven items total were analyzed which included the original 31 School Implementation Scale items as it was created with an additional 6 items regarding other state initiatives being implemented (Collaborative Data Teams and Common Formative Assessments) as requested by the state. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.949, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and a Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 62068.964 (666 degrees of freedom and $p<0.001$). Seven factors were extracted with Eigenvalues larger than 1.0, and accounted for a cumulative 62% of the variance. Table 34 shows the Eigenvalues and percent of variance for each of the 7 extracted factors- one very large factor with an Eigenvalue of 12.9 and accounting for 35% of variance; a smaller yet prominent factor with an Eigenvalue of 3.2 and accounting for almost 9% of the variance; 2 smaller factors with Eigenvalues of 1.8 (almost 5%) and 1.6 (4%); and 3 much smaller factors with Eigenvalues of 1.2 (3%), 1.1 (3%), and 1.1 (almost 3%).

Table 34. Midwestern CW 37 Items- EFA Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	12.880	34.810	34.810
2	3.148	8.509	43.319
3	1.753	4.738	48.056
4	1.602	4.329	52.385
5	1.197	3.235	55.620
6	1.131	3.058	58.679
7	1.067	2.885	61.564

Table 35 displays a summary of the rotated item-factor structure resulting from the SPSS EFA for the Midwestern CW 37 item version of the instrument as administered, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 11. Of the 13 items theorized to be in the SIS factor of Evidence-Based Practices, 9 loaded on Factor #1 with 6 excellent, 2 very good, and 1 good item-loading values according to Comrey & Lee (1992). Two theorized Evidence-Based Practices items had excellent item-factor loadings on Factor #6, and 2 theorized Evidence-Based Practices items had fair and very good item-factor loadings on Factor #7.

Table 35. Midwestern CW 37 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	13	#1	6	2	1		
		#6	2				
		#7		1		1	
School Culture	9	#2	1	3	4		
		#1			1		
Family Involvement	5	#4	3				
		#1			1		
		#2		1			
Professional Development	4	#5	3	1			
Additional State Items	6	#3	1	2	3		

Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #2 with 1 excellent, 3 very good, and 4 good item-factor loadings; and the final School Culture item showed a good item-loading value on Factor #1. Of the 5 items theorized to be included in the Family Involvement SIS factor items, 3 items showed excellent item-factor loadings on Factor #4. One Family Involvement theorized item had a good loading on Factor #1, and the final Family Involvement item had a very good item-factor loading on Factor # 2. Of the 4 items theorized to be included in the Professional Development SIS factor, all loaded on Factor #5 with

3 excellent and 1 very good item-factor loadings. Of the 6 state-requested additional items, all 6 items loaded on Factor #3 with 1 excellent, 2 very good, and 3 good item loading values.

Western State Effective Behavioral & Instructional Support Systems with 42 Items- Exploratory Factor Analysis without specifying the number of factors to extract was conducted in SPSS on the 2013-2014 Western state's complete dataset including all survey items as administered. Forty-two items total were analyzed which included 28 of the original 31 School Implementation Scale items as it was created, and an additional 14 items regarding other state initiatives being implemented (13 pertaining to tiered levels of academic and behavior supports and 1 pertaining to communicating with families about student progress toward meeting Common Core Standards) as requested by the state. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.949, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and a Bartlett's Test of Sphericity rendered an Approximate Chi Square value of 39168.755 (861 degrees of freedom and $p<0.001$). Seven factors were extracted with Eigenvalues larger than 1.0 and accounted for a cumulative 64% of the variance.

Table 36 shows the Eigenvalues and percent of variance for each of the 7 extracted factors- one very large factor with an Eigenvalue of 15.7 and accounting for 37% of variance; a smaller yet prominent factor with an Eigenvalue of 4.5 and accounting for almost 11% of the variance; then 2 smaller factors with Eigenvalues of 1.7 (4%) and 1.5 (almost 4%); and 3 much smaller factors with Eigenvalues of 1.4 (3%), 1.2 (almost 3%), and 1.0 (2%).

Table 36. Western EBISS 42 Items- EFA Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	15.651	37.264	37.264
2	4.516	10.752	48.016
3	1.734	4.128	52.144
4	1.470	3.501	55.645
5	1.426	3.396	59.041
6	1.168	2.780	61.821
7	1.017	2.422	64.243

Table 37 displays a summary of the rotated item-factor structure resulting from the SPSS EFA for the Western EBISS 42-item version of the School Implementation Scale as administered, and the full Rotated Factor Structure Matrix can be found in Appendix E Table 12. Of the 10 items theorized to be in the SIS factor of Evidence-Based Practices, 4 loaded on Factor #6 with 2 excellent and 2 very good item-loading values according to Comrey & Lee (1992). Two theorized Evidence-Based Practices items had excellent item-factor loadings and 1 item had a good loading on Factor #3, two theorized Evidence-Based Practices items had excellent item-factor loadings on Factor #2, and 1 item had a very good loading on Factor #5.

Five of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #5 with 1 excellent, 3 very good, and 1 poor item-factor loadings. Two theorized School Culture items showed good and poor item-loadings value on Factor #1, 1 item had a very good loading on Factor # 3, and the final theorized School Culture item had a fair item-factor loading on Factor #3. Of the 5 items theorized to be included in the Family Involvement factor, 3 items showed excellent item-factor loadings on Factor #4. One Family Involvement theorized item had a good loading on Factor #1, and the final Family Involvement item had a good loading on Factor #6. Of the 4 items theorized to be included in the SIS Professional Development factor, 2 loaded on Factor #2 with excellent and good item-factor loadings. One theorized Professional

Development item showed a good loading on Factor #6, and the final theorized Professional Development item showed an excellent loading on Factor #7.

Table 37. Western EBISS 42 Items- EFA Item Loading Summary

			Item-Factor Loading Values				
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor
Evidence-Based Practices	10	#6	2	2			
		#3	2		1		
		#2	2				
		#5		1			
School Culture	9	#5	1	3			1
		#1			1		1
		#3		1		1	
Family Involvement	5	#4	3				
		#1			1		
		#6			1		
Professional Development	4	#2	1		1		
		#6			1		
		#7	1				
Additional State Items	13	#1	4	2			
		#3	3	1			
		#5		2			
		#7	1				

Of the 13 state-requested additional items pertaining to tiered levels of support, 6 items loaded on Factor #1 with 4 excellent and 2 very good item loading values. Four of the additional tiered levels of support items loaded on Factor #3 with 3 excellent and 1 very good item-factor loadings. Two state-requested tiered levels of support items showed very good loadings on Factor #5, and the final tiered levels of support item showed an excellent loading on Factor #7. The 1 additional state-requested item regarding communicating with family about student progress toward meeting Common Core Standards showed an excellent item-factor loading on Factor #4.

CHAPTER 4

Conclusions

This study explores the validity and reliability of the School Implementation Scale, and its' administration across various state populations (the Midwestern state's Collaborative Work initiative, the Western state's Effective Behavioral and Instructional Support Systems initiative, and both states combined) and different forms of multi-tiered support systems (School-Wide Positive Behavior Support from the Midwestern state with Schoolwide Positive Behavior Intervention Supports from the Western state, and Response-to-Intervention from the Western state). The statistical analysis of the five datasets which consisted of the 28 School Implementation Scale items that were common between the two states from the original 31 scale reveals that the 4-factor structure specified for the SIS does not adequately fit any of the datasets examined.

The confirmatory factor analysis conducted in Mplus on the five datasets for all of the populations examined- Midwestern CW, Western EBISS, SW-PBS/SWPBIS, RTI, and the two state datasets combined- reveal values for the goodness of fit tests that indicate the 4-factor structure specified for the School Implementation Scale does not adequately fit any of the datasets. The lack of adequate model fit for the 2014 data of both states combined which consisted of the 28 common SIS items across the two states means there is not a strong base model from which to examine the measurement invariance between the overall education initiatives of the Midwestern and Western states (research question 2) or between the different multi-tiered systems of support programs of SW-PBS/SWPBIS and RTI (research question 3), and the remaining statistical analyses planned for the study could not be conducted. Due to the fact that the measurement invariance of the SIS factor structure could not be tested, and thus, does not hold to any extent; the reliability of results across the two states or the two different RTI programs in terms of respondent demographic characteristics (research question 4) could not be investigated, and the differences between and patterns within respondents' mean SIS and mean essential element domain scores in terms of respondent demographic characteristics (research question 5) could not be examined.

The exploratory factor analysis specifying 4 factors conducted in SPSS reveals that the percentage of variance in the data accounted for by the four extracted factors is under 60% for each the five datasets with a range from 56.5% to 58.8%. An oblique rotation procedure was employed to ease the interpretation of the item-factor loadings but many items show low item-factor loadings on their intended factor, cross-load with adequate loading values on multiple factors, and/or fail to achieve an interpretable item-factor loading on their intended factors. The reliability estimates reveal fairly high values for internal consistency of items for the overall 28-item SIS (0.937-0.945) and for the four individual factors (Evidence-Based Practices 0.865-0.893, School Culture 0.818-0.851, Family Involvement 0.797-0.832, and Professional Development 0.805-0.829), but there are items that if deleted would increase either the Cronbach's Alpha for the overall School Implementation Scale and/or the essential elements of successful school factors supposedly measured by the scale. This finding indicates that some of the 28 common SIS items either do not fit well with the other items theorized to be included in the factor or in the overall scale, that some items are not very discriminating for their intended factor or the overall scale, or that items are measuring constructs that are extremely similar to what other items are measuring.

The exploratory factor analysis without specifying a number of factors to be extracted that was conducted in SPSS shows the number of factors extracted from the different datasets varied from 4 to 6, and that many items load on factors other than the factor for which inclusion of the item was intended by the item-factor structure outlined.

However, the 4-factor item-structure specified for the School Implementation Scale was originally created by Gaumer Erickson, et al. (2012) with 31 items. The 2014 Midwestern state's Collaborative Work dataset contains the original 31 School Implementation Scale instrument items. The 31 SIS items included 13 items in the construct of Effective Practices, 9 items under the construct of School Culture, 5 items in the Family Involvement construct, and 4 items under the Professional Development construct as the theorized 4-factor structure of the SIS outlined. Confirmatory Factor Analysis (CFA) was conducted in Mplus on the 2014 Midwestern Collaborative Work state-wide initiative data which consisted of the original 31 SIS items, and the 4-factor structure model outlined for the School Implementation Scale was specified. Model fit indices were interpreted using the acceptable fit guideline values from a review of reporting CFA results (Schreiber, et al., 2006). The resulting model fit statistics rendered a Comparative Fit Index (CFI) equal to 0.800, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.783, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal 0.090, but a value of 0.06 or less means acceptable model fit (Steigler, 2007). CFA results for the Midwestern state 31-item data rendered a significant Chi Square test of base model fit value of 54718.766 with 465 degrees of freedom and $p<0.001$, but a non-significant p-value and low Chi Square value relative to the degrees of freedom are considered acceptable for good model fit (Hooper, et al., 2008). Thus, the 2014 Midwestern 31-item data did not show adequate model fit to the 4-factor item-structure specified for the School Implementation Scale.

The resulting Standardized Parameter Estimates, Factor Score Coefficients, and Covariance Correlations for the Mplus CFA on the Midwestern CW 31-item data can be found in Appendix D Tables 16, 17, and 18. The standardized parameter estimates, or item-factor loading coefficients, for Factor #1 revealed a range from 0.333 to 0.824. The items with the lowest coefficients were C3, C8, C1, and C5 with estimates less than 0.600; but items C2, C6, C7, C4, and C9 showed standardized parameter estimates of 0.700 or larger. For Factor #2 the item-factor loadings ranged from 0.385 to 0.785. Items E1, E13, E2, E8, E12, E7, and E10 showed standardized parameter estimates lower than 0.500; and items E4, E6, E5, E3, E11, and E9 had item-factor loadings larger than 0.600. Factor #3 showed item-factor loadings ranging from 0.513 to 0.810, and while item P2 had a coefficient value around 0.500, the other 4 items had coefficient values larger than 0.700. For Factor #4 the standardized parameter estimates ranged from 0.363 to 0.870. Item F1 showed an item-factor loading values less than 0.400; items F4 and F5 show standardized parameter estimates between 0.500 and 0.700; and items F4 and F2 showed item-loading values larger than 0.800.

Factor score coefficients for the Midwestern CW 31-item CFA in Mplus ranged from 0.041 to 0.099 for the theorized School Culture items on Factor #1, and were lower for all other factors (see Appendix D Table 17). The theorized Evidence-Based Practices items revealed factor score coefficients ranging from 0.019 to 0.062 on Factor #2, and the Professional Development items revealed values ranging from 0.111 to 0.232 on Factor #3. Factor #4 showed factor score coefficients ranging from 0.014 to 0.164 for the theorized Family Involvement items.

The finding that the 2014 Midwestern state's Collaborative Work data with all 31 of the original scale items does not fit the 4-factor structure specified for the SIS was cross-validated with the 2013 Midwestern CW data consisting of the original 31 SIS items. A confirmatory factor analysis was conducted in Mplus on the 2013 Midwestern CW 31-item data. The Comparative Fit Index (CFI) was equal to 0.786, but a value of 0.95 or greater means an acceptable model fit. The Tucker-Lewis Fit Index (TLI) was equal to 0.767, but a value of 0.95 or greater means an acceptable model fit. The Root Mean Square Error of Approximation (RMSEA) was equal to 0.095, but a value of 0.06 or less means acceptable model fit (Steigler, 2007). CFA results for the 2013 Midwestern state data rendered a significant Chi Square test of base model fit value of 45314.677 with 465 degrees of freedom and $p<0.001$. These results show that, as with the 2014 data, the 4-factor item-structure model specified by the SIS does not adequately fit the 2013 Midwestern state's CW initiative 31-item data.

Exploratory factor analysis specifying 4 factors be extracted and employing Principle Axis Factoring and Promax Rotation was conducted in SPSS on the 2014 Midwestern CW data including the 31 SIS items to examine whether the original scale collects data that fit the 4-factor structure specified by the SIS. The theorized 4-factor structure accounted for a cumulative 56% of the variance. Table 38 shows the Eigenvalues and percent of variance for each of the 4 extracted factors- one very large factor with an Eigenvalue of 12.2 accounting for 39% of the variance; a smaller yet prominent factor with an Eigenvalue of 2.3 accounting for 7% of the variance; and 2 much smaller factors with Eigenvalues of 1.7 (almost 6%) and 1.2 (4%).

Table 38. Midwestern CW 31 Items- EFA specifying 4 factors Total Variance Explained

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	12.203	39.365	39.365
2	2.312	7.458	46.823
3	1.720	5.549	52.372
4	1.217	3.926	56.299

Table 39 displays a summary of the rotated item-factor structure for the Midwestern state's Collaborative Work 31-item version of the SIS instrument resulting from the SPSS EFA specifying 4 factors. The full Rotated Factor Structure Matrix can be found in Appendix E Table 13. Of the 13 items theorized to be in the SIS factor of Evidence-Based Practices, 7 items loaded on Factor #2 with 2 excellent, 2 very good, 2 good, and 1 fair item-factor loading values. Five of the theorized Evidence-Based Practices items loaded on Factor #3 with 3 excellent and 2 very good item-loading values, and showed good and poor cross-loadings on Factor #2; and the final theorized Evidence-Based Practices item loaded on Factor #4 with a good item-loading value, and showed a poor cross-loading on Factor #2. Cronbach's Alpha for the Evidence-Based Practices factor was 0.904, and the deletion of item E4 (I review universal screening data at least three times a year for every student that I support) would increase the internal consistency measure for the factor to 0.905.

Table 39. Midwestern CW 31 Items- EFA specifying 4 factors Item Loading Summary

			Item-Factor Loading Values							
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load	Cross Loadings	Cronbach's Alpha
Evidence-Based Practices	13	#2	2	2	2	1				0.904
		#3	3	2					5 #2	
		#4			1				1 #2	
School Culture	9	#1	1	3	3	1				0.851
		#3			1				1 #1	
Family Involvement	5	#4	3							0.832
		#1		1					1 #4	
		#2			1				1 #4	
Professional Development	4	#3	1	1	1					0.823
		#1		1					1 #3	
Total SIS	31									0.945

Eight of the 9 items theorized to be in the SIS factor of School Culture loaded on Factor #1 with 1 excellent, 3 very good, 3 good, and 1 fair item-factor loadings; and the final School Culture item showed a good item-loading value on Factor #3 with a poor cross-loading value on

Factor #1. Cronbach's Alpha for the School Culture factor was 0.851, and the elimination of item C3 (I have a clear understanding of the State Standards for my grade/subject) would increase the internal consistency measure for the factor to 0.853.

Of the 5 items theorized to be included in the Family Involvement SIS factor, 3 items showed excellent item-factor loadings on Factor #4. One Family Involvement theorized item had a very good loading on Factor #1 with a fair cross-loading on Factor #4, and the final Family Involvement item had a good item-factor loading value on Factor #2 with a fair cross-loading on Factor #4. Cronbach's Alpha for the Family Involvement factor was 0.832. However, the elimination of item F1 (I consider my students' backgrounds when planning instruction) would increase the internal consistency measure for the factor to 0.835, and the elimination of item F5 (I think my school does a good job of including parents as team members in data-based decision making) would increase Cronbach's alpha for the Family Involvement factor to 0.844.

Of the 4 items theorized to be included in the Professional Development factor, 3 loaded on Factor #3 with 1 excellent, 1 very good, and 1 good item-factor loading values; and the final theorized Professional Development item had a very good item-factor loading on Factor #1 with a fair cross-loading on Factor #3. Cronbach's Alpha for the Professional Development factor was 0.823, and the elimination of any of the 4 items did not increase the internal consistency measure for the factor.

The Cronbach's Alpha for the 2014 Midwestern CW 31-item School Implementation Scale was 0.945, which showed a high level of internal consistency among the 31 SIS items. The full Item-Statistics table for the School Implementation Scale, including the Alpha if item deleted, can be found in Appendix F Table 6. It was revealed that the elimination of item C5 (I have the technology and resources that I need to provide effective instruction) would increase the internal consistency measure for the scale to 0.946. The alpha decreased slightly to 0.944 if items C1, C3, C6, C8, E1, E2, E4, E7, E8, E10, E12, E13, P2, F1, F2, F3, or F5 were deleted.

The rotated item-factor structure of the SPSS EFA specifying 4 factors for the 2014 Midwestern CW 31-item dataset seemed to suggest that Factor #1 was School Culture, Factor #2 was Evidence-Based Practices, Factor #3 was Professional Development, and Factor #4 was Family Involvement. The Factor Correlation Matrix in Table 40 shows that the highest correlated factors were School Culture and Professional Development at 0.614, followed by Evidence-Based Practices and Professional Development at 0.609, and Family Involvement and Professional Development at 0.565. Family Involvement and Evidence-Based Practices at 0.533, School Culture and Family Involvement were correlated at 0.523, and the lowest correlation between factors was School Culture and Evidence-Based Practices at 0.477.

Table 40. Midwestern CW 31 Items- EFA specifying 4 factors Factor Correlation Matrix

	School Culture	Evidence-Based Practices	Professional Development	Family Involvement
School Culture	1.000	.477	.614	.523
Evidence-Based Practices	.477	1.000	.609	.533
Professional Development	.614	.609	1.000	.565
Family Involvement	.523	.533	.565	1.000

An exploratory factor analysis without specifying a number of factors to be extracted was conducted in SPSS on the 2014 Midwestern state’s Collaborative Work dataset consisting of the original 31 School Implementation Scale instrument items. The examination of factorability showed a Kaiser-Meyer-Olkin Measure of Sampling Adequacy equal to 0.952, above the 0.6 value required for good factor analysis (Tabachnick & Fidell, 2006, p.614) and approaching 1.0 meaning that the partial correlations between items were small, and the Bartlett’s Test of Sphericity rendered an Approximate Chi Square value of 54606.000 with 465 degrees of freedom and p<0.001. Six factors were extracted with Eigenvalues greater than 1.0 and accounted for 63% of the variance. The factors extracted included a very large factor with an Eigenvalue of 12.2 accounting for 39% of the variance; two smaller yet prominent factors with Eigenvalues of 2.3 and 1.7 accounting for 7% and almost 6% of the variance, respectively; and two factors with Eigenvalues of 1.08 (3.5%) and 1.05 (3.4%).

Table 41 displays a summary of the rotated item-factor structure for the Midwestern state's Collaborative Work 31-item version of the SIS instrument resulting from the SPSS EFA. The full Rotated Factor Structure Matrix can be found in Appendix E Table 14. Of the 13 items theorized to be in the SIS factor of Evidence-Based Practices, 8 items loaded on Factor #1 with 3 excellent, 4 very good, and 1 good item-factor loadings. Two of the theorized Evidence-Based Practices items loaded on Factor #5 with excellent item-loading values, and 3 theorized Evidence-Based Practices items loaded on Factor #6 with 2 very good and 1 good item-loading values.

Table 41. Midwestern CW 31 Items- EFA Item Loading Summary

			Item-Factor Loading Values					
Theoretical Factor	Num Items	Extracted Factor	Excellent	Very Good	Good	Fair	Poor	Did Not Load
Evidence-Based Practices	13	#1	3	4	1			
		#5	2					
		#6		2	1			
School Culture	9	#2	1	2	4			
		#1		1				
		#6		1				
Family Involvement	5	#3	3		1			
		#1			1			
		#2		2				
Professional Development	4	#4	3	1				
Total SIS	31							

Thus, it seems that the item-factor structure proposed by Gaumer Erickson, et al. (2012) designating the four latent domains of School Culture, Evidence-Based Practices, Professional

Development, and Family Engagement may be an incorrectly specified model for the SIS. This study concludes that the School Implementation Scale with its current items-factor structure for either 28 or 31 items is not a valid and reliable measure of the implementation level of integrated academic and behavioral multi-tiered systems, and the use of the SIS across various state populations and different forms of multi-tiered support systems would not be recommended.

CHAPTER 5

Discussion

Gaumer Erickson, et al. suggested in 2012 that future research regarding the School Implementation Scale should include administration in schools implementing various integrated academic and behavior tiered RTI models in order to examine the performance of the SIS over different populations, overall educational initiatives, and different integrated academic and behavior tiered RTI models. With the 2013-2014 data collected from two states and consisting of multiple integrated academic and behavior tiered RTI models, these research needs could be addressed; and by expanding the previous validation study methodology through the use of more advanced statistical analysis, the existing reliability and validity evidence related to the School Implementation Scale (SIS) could be enhanced. This additional reliability and validity evidence would better inform evaluators and educators to accurately measure the implementation of integrated academic and behavior RTI models of tiered supports using the essential elements of successful schools theoretically measured by the SIS.

However, the supplementary psychometric information resulting from this study of the SIS does not justify the continued administration of the instrument for evaluation research, and the measure currently lacks utility in comparing various integrated academic and behavior tiered RTI models in terms of measuring and monitoring implementation. Therefore, educators and evaluators should be cautioned against making decisions based on the School Implementation Scale or its four essential elements of successful school domain scores for School Culture, Evidence-Based Practices, Professional Development, and Family Engagement that may not be reliable or valid.

All scales, especially those used in educational evaluation, should be developed according to, and examined extensively following, the guidance of The Standards for Psychological and Educational Testing (2014), The Program Evaluation Standards (2011), and the AEA Guiding Principles for Evaluators (2004). Not all measurement instruments meet the professional standards for program evaluation or educational and psychological testing. "Adequate measures are a necessary condition for valid research," and "poor measurement imposes an absolute limit on the validity of the conclusions one can reach" (DeVellis, 2003, p.12). As shown by the results of the statistical analyses conducted for this study, there are major psychometric issues associated with the theorized item-factor structure specified for the School Implementation Scale by Gaumer Erickson, et al. (2012) and the lack of model fit to the data collected with the measure in the Midwestern and Western states during 2013 and 2014.

Limitations

There could be multiple reasons that the 4-factor structure of the SIS does not fit the datasets. One possibility is the condition of the data itself. The data collected with the SIS tends to exhibit negative skewness, kurtosis, incidents of item multicollinearity, large standard deviations and error variance, and differing response patterns among groups of respondents. According to Reise, Waller, & Comrey (2000), sample heterogeneity can also cause an issue with identifying replicable factors because samples with "sufficient examinee representation at all levels of the trait dimensions" are required "in order to accurately estimate the population inter-correlations" (p. 290). These issues, as well as the acquiescence bias of self-reported data (Welkenhuysen-Gybels, Billiet, & Cambre, 2003) in which respondents have a tendency to answer questions positively or confirm the item statements, can affect survey responses and could have contributed to the lack of model fit to the data.

Another form of response bias, social desirability bias, could also have played a role contributing to the negative skewness and inadequate fit of the item-factor structure to the data. Social desirability bias is the tendency for a person to answer questions on a survey untruthfully due to feeling pressure of providing responses that are socially acceptable. This leads to a predisposition for survey respondents to answer questions in a manner that will be viewed favorably by others either by over-reporting desirable behaviors or under-reporting behaviors deemed undesirable. The pressure to achieve this social desirability, and the potential consequences of not gaining others' approval could definitely impact the way school staff respond to the items on the instrument. Individual educators might report higher levels of implementation than reflected in their everyday practices, and school administrators or instructional coaches associated with the educational initiative might indicate that the district/state model is more successful than in reality so as to keep funding and resources.

Fidelity of implementation, also referred to as treatment integrity, is the degree to which programs are implemented as intended by the program developers. Fidelity and "consistency in implementation throughout the school is an assumption often made by RTI initiatives" (Gaumer Erickson, et al., 2012, p.36). Although, "in reality, there are many situational factors- inside and outside the classroom- that support and account for [the RTI's] successful implementation" (Fuchs & Deschler, 2007, p.131). Through a state-wide initiative Michigan launched a multitiered system of supports as a framework to improve academic and behavior in more than half of the state's schools and found that initial buy-in from staff, financial incentives, or even early success does not guarantee schools will be able to sustain the model (Sparks, 2016). The executive director of curriculum and staff development for one Michigan school district, Steve Netzel, maintains that the model "is a recipe, it's not a McDonald's value menu" (Sparks, 2016, p.10) from which schools can choose which parts of the multitiered model they like and want to implement. Netzel continues by stating that "it takes a while to understand it's a system and it all interacts with each other. You can't pick and choose" (Sparks, 2016, p.10) or implementation would vary significantly from school to school. Reinke, Herman, & Stormont (2013) found that school-wide positive behavior intervention models will not be as effective in supporting outcomes for students if ineffective management practices are present at the classroom level, but that some dimensions of SW-PBIS are more easily incorporated into the classroom without further training or professional development than are other dimensions of the model.

Dulaney, Hallam, & Wall conducted a descriptive case study in 2013 examining the perceptions of district superintendents regarding opportunities and obstacles associated with Multi-tiered Systems of Support (MTSS) implementation. Analysis of the data showed that superintendents believe that districts must develop the MTSS framework and promote a common language based on the framework, that there must be a district-wide culture of collaboration, and that the capacities of individuals must be built at every system level so improvement is ongoing and sustainable (Dulaney, et al., 2013). A case study of the experiences of Florida's Positive Behavior Support Project conducted by Kincaid, Childs, Blasé, & Wallace in 2007 used a "systematic process to understand barriers and facilitators to the successful implementation of schoolwide positive behavior support by schools implementing at high and low levels of fidelity" (p.174). Results indicated that schools implementing with low fidelity tend to identify practical, operational barriers, whereas schools implementing at high fidelity struggle with systems issues. Most of the thirteen common themes identified having to do with barriers and facilitators to implementation themes reflect the core components for initiating and maintaining SWPBS; including obtaining administrative and district support developing a reward system for students and staff, obtaining staff buy-in, using data, working as a team, and involving families and the community (Kincaid, et al., 2007). Implementation fidelity is an important source of variation affecting the credibility and utility of research, and “it has been demonstrated that the fidelity with which an intervention is implemented affects how well it succeeds” (Carroll, Patterson, Wood, Booth, Rick, & Balain, 2007). If there is a lack of fidelity or consistency in the implementation of the integrated academic and behavioral multi-tiered systems of support, the School Implementation Scale responses could be inconsistent making it difficult to separate items into well-defined latent constructs, or respondents could be artificially inflating their answers in attempt to over-compensate for the lower levels of implementation fidelity.

The major limitation of this study is that, due to the lack of model fit between the School Implementation Scale the theorized 4-factor item-structure and the data, much of the statistical analyses planned for the study could not be conducted. Research questions left unexamined include the extent of measurement invariance of the latent constructs measured by the SIS within and between states with different educational initiatives and integrated academic and behavior tiered response-to-intervention models; the ability of the instrument overall and its four essential elements of effective schools domains to provide reliable data for all school staff across respondent’s role, years of experience, and involvement with the leadership team; and what differences and patterns might be observed in respondents’ mean SIS and mean essential element domain results in terms of staff role, years at the school, involvement with the leadership team, or the interactions of these demographics variables. These research questions cannot be addressed until the underlying item-factor structure of the School Implementation Scale, and its proposed four latent constructs or essential elements of successful schools domains, are reviewed and revised.

Future Recommendations

According to Reise, Waller, & Comrey (2000) the goal of scale revision is to improve the psychometric properties and the validity of the measurement instrument. Improving the validity of the instrument would mean assuring that the measure has item content and a corresponding factor structure that is representative of and consistent with what is currently known regarding a construct, and identifying a factor structure that is replicable and generalizable across relevant populations. “The primary motivations for scale revision is that the scale’s psychometric properties are deemed inadequate, that research may demonstrate that an existing measure has a factor structure that is not generalizable across different samples or has a factor structure that is ‘not as advertised’ by the original authors, and/or there is inadequate content representation” (Reise, et al., 2000, p.288). Additionally, to inform and update the content represented in the items, qualitative data should be collected via interviews and focus groups with educators that have experience with the implementation of integrated academic and behavioral multi-tiered systems of support in their schools. These individuals’ experience could prove vital in identifying the barriers and facilitators of successful implementation and sustainability of the core components or essential elements of the educational model.

A summary of fundamental psychometric criteria that should be considered at all stages of the scale revision process was provided by Smith & McCarthy in 1995. These criteria include (a) recognizing a scale’s hierarchical structure (i.e., what facets of item content it contains), (b) establishing internal consistency reliability when appropriate, (c) testing of content homogeneity of the facets and ensuring that different aspects of the construct are equally represented in a scale, (d) ensuring that they items discriminate between respondents at the appropriate level of trait intensity, and (e) replication of the factor structure across independent samples (Smith & McCarthy, 1995). Reise, Waller, and Comrey (2000, p. 288) describe two guiding principles for a researcher planning to use EFA to develop, refine, and evaluate a measure. The first guiding principle is that the researcher should develop a clearly articulated plan regarding the needs for the revised instrument- addressing the questions of what construct is being measured, why does the construct need measured, what level of the construct hierarchy is the measure, and how the measure is different from other competing measures. The second guiding principle states that the scale developer should conduct a systematic series of studies using large samples of respondents, and that factor-analytic-based scale revision should be an iterative process where data inform construct definition and refinement.

The first recommended step in revising the School Implementation Scale would be eliminating the items from the overall instrument and from each of the essential elements of successful school domains or latent factors that, when deleted, increase the Cronbach's alpha measure of internal consistency. This would consist of deleting item C5- "I have the technology and resources that I need to provide effective instruction" from the overall scale which increases the alpha from 0.945 to 0.946. Item C3- "I have a clear understanding of the state standards for my grade/subject" would be deleted from the School Culture domain increasing Cronbach's alpha for the domain from 0.851 to 0.853, and item E4- "I review universal screening data at least three times a year for every student that I support" would be deleted from the Evidence-Based Practices domain increasing Cronbach's alpha for the domain from 0.904 to 0.905. Two items would be deleted from the Family Engagement domain: Item F1- “I consider my students' backgrounds when planning instruction” increasing Cronbach's alpha from 0.832 to 0.835, and item F5- “I think my school does a good job of including parents as team members in data-based decision making” increasing Cronbach's alpha from 0.832 for the domain to 0.844. No items would be removed from the Professional Development domain, as there were no items that if deleted would improve the internal consistency of the domain.

The next recommended step in revising the School Implementation Scale would be combining items that are so highly correlated with each other, that the separate items are practically measuring the same construct. Items E12- “When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions” and E13- “When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions” have a Pearson bivariate correlation of 0.738. It is suggested that these two items become one Evidence-Based Practices domain item that states “When I'm concerned about a student's academic or behavioral progress, I collaborate with colleagues to identify interventions.” Items F2- “I regularly communicate with families regarding student academic goals/progress” and F3- “I regularly communicate with families regarding student behavioral goals/progress” have a Pearson bivariate correlation of 0.870. It is suggested that these two items become one Family Engagement domain item that states “I regularly communicate with families regarding student academic and behavioral goals/progress.”

The procedures outlined above would decrease the number of items on the School Implementation Scale from 31 to 24. This would also impact the number of items included the latent factors and further shrink the content coverage for the elements of successful school domains of School Culture (from 9 items to 7), Evidence-Based Practices (from 13 items to 11), and Family Engagement (5 items to 2); but the Professional Development domain would not change in terms of the number of items. The number of items per theorized latent construct vary greatly with Evidence-Based Practices having the greatest number of items, and therefore more content associated with evidence-based practices is represented on the overall scale. The domain of School Culture with 7 items is probably covered sufficiently to define the construct, but the domains of Professional Development with 4 items and Family Engagement with 2 items do not probably contain enough content associated with the constructs of those domains to be properly defined. Either more items should be created to improve the content coverage and help define the Professional Development and Family Engagement constructs being measured, or these items should be re-categorized to combine with other remaining latent constructs that are similar or closely related to the items' previous domains.

Considering the number of items that would now be included in each of the essential element of successful schools domains, and the correlations between the latent constructs as previously defined, it is suggested that a two-factor item-structure might be further investigated for the School Implementation Scale. Combining the Evidence-Based Practices and Professional Development constructs would lead to a total of 15 items, and combining the School Culture and Family Engagement constructs would lead to a total of 9 items. The increased number of items per domain should help to better define the constructs being measured due to the enhanced coverage of the content that previous research has shown should be included in the domain. Given the high correlation between these two pairs of domains, it follows that the items would yield a fairly high level of internal consistency within those newly-defined latent constructs.

Ongoing professional development is an evidence-base practice that informs and trains educators to implement new instructional strategies and data-driven decision making into their every-day practices; thus, these domains and their associated items should produce a more robust and valid Evidence-Based Practices construct. The extent to which families are included as partners in the education of students is mostly due to an environment where their involvement is accepted, encouraged, and valued; but this can only be accomplished through the collaborative culture of the school. Including the family engagement items with the items examining other aspects of the school culture should also produce a more robust and valid School Culture construct.

The revision the School Implementation Scale must be an iterative process in which after deleting items, reliability analysis should be conducted to assure an acceptable level of Cronbach's alpha for the overall School Implementation Scale and for each of the measured domains separately. Bivariate Pearson correlations between all SIS items, between the items within each domain, and the factor correlations between the domains should be checked throughout the revision process and will help to guide further edits regarding item inclusion or combining items. Exploratory factor analysis employing an oblique rotation technique should be conducted after item changes are made, and the Eigenvalues of the extracted factors or the percentage of overall variance explained by the factors should be examined to assure that each factor contributes substantially to the structure model. The rotated factor structure matrix resulting from the EFA will then need to be assessed for reasonable item-factor loadings are obtained for each item on their intended latent constructs and negligible cross-loadings on other factors exist. The suggested two-factor structure for the SIS should then be subjected to confirmatory factor analysis using Mplus to test for adequate model fit to the data collected with the newly revised measure.

Once the School Implementation Scale has undergone the iterative scale revision process and an alternative item-factor structure has been identified that achieves an acceptable model fit for the data, the remaining research questions posed by this study should be conducted before the instrument can be considered a reliable and valid measure for use in educational evaluation. Statistical analysis pertaining to the extent of measurement invariance of the revised item-factor structure model over varying integrated academic and behavioral multi-tiered systems of support, across distinct populations, and for differing respondent demographic characteristics will need to be conducted for the revised SIS. However, it is recommended that future research give additional consideration to employing procedural and statistical remedies that minimize the detrimental effects of common method biases which can influence the estimates of construct validity and reliability and influence the covariation between different constructs (Podsakoff, MacKenzie & Podsakoff, 2012).

It should also be noted that the data collected using the SIS, when combined as in this study, is naturally clustered and exhibits complex patterns of variability. Responses are nested within school, within district, and within state; as well as across time if respondents complete the instrument more than once a year or if years of data containing the same respondents are combined. None of these clustered categories are randomly assigned or independently sampled; therefore, it is suggested that hierarchical statistical methods (Raudenbush & Bryk, 2002) be employed when conducting analysis that would include any kind of group comparisons.

In conclusion, it is highly recommended that the School Implementation Scale be revised and its psychometric properties analyzed as Gaumer Erickson, et al.'s 2012 statement still holds true- "it is necessary to develop common evaluation measures that can inform data-based decision making within schools, evaluate the outcomes of multiple models, and compare models to provide a framework for continued improvement and facilitate the cross-flow of information related to effective practices" (p.49) in order to successfully implement integrated academic and behavioral multi-tiered systems of support and improve the effectiveness of educators and the outcomes for students.

References

Algozzine, B., Wang, C., and Violette, A.S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions*, 13, 3-16.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

American Evaluation Association (2004). G*uiding Principles for Evaluators*. Retrieved from, http://www.eval.org/p/cm/ld/fid=51.

Barrett, P. (2007). Structural equation modeling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Bland, J.M. & Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *Bmj, 310*(6973), 170.

Borsboom D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425-440.

Borsboom, D., Mellenbergh, G.J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071.

Bowen, N. K. (2014). Testing for differences in measurement (CFA) models using Mplus's invariance shortcut code (WLSMV). *Unpublished manuscript*.

Bradley, R., Danielson, L., & Doolittle, J. (2007). Responsiveness to intervention: 1997 to 2007. *Teaching Exceptional Children*, *39*(5), 8-12.

Bradshaw, C. P., Debnam, K., Koth, C. W., & Leaf, P. (2009). Preliminary validation of the implementation phases inventory for assessing fidelity of schoolwide positive behavior supports. *Journal of Positive Behavior Interventions, 11*(3), 145-160.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A. & Long, J.S. (eds.), *Testing Structural Equation Models*, 136-162. Newbury Park, CA: Sage Publications.

Buss, A. R., & Royce, J. R. (1975). Detecting cross-cultural commonalities and differences: Intergroup factor analysis. *Psychological Bulletin*, *82*(1), 128.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological bulletin*, *105*(3), 456.

Carroll, C., Patterson, M., Wood, S. Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 40, 1-9.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, *14*(3), 464-504.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-235.

Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring school-wide positive behavior support implementation development and validation of the benchmarks of quality. *Journal of Positive Behavior Interventions*, *9*(4), 203-213.

Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*(1), 180-213.

Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research*, *20*, 389-417.

Comrey, A.L. & Lee, H.B. (1992). *A First Course in Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, D.A. & Beckman, T.J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine, 119*(2), 166.37-166.e16.

Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* SAGE Publications, Inc.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Holt, Rinehart and Winston.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. P*sychometrika*, *16*(3), 297-334.

Cronbach, L. J., & Thorndike, R. L. (1971). Educational measurement. *Test validation*, 443-507.

Cureton, E. E. Validity. (1951). In Lindquist, E. F. (Ed). *Educational measurement. American Council on Education, 19*, 621-694.

DeVellis, R. F. (2003). *Scale Development: Theory and Application (2nd ed.)*. Thousand Oaks, CA: Sage Publications, Inc.

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are central issues. *Psychological Bulletin*, *95*, 135-135.

Drasgow, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, *72*, 19-29.

Drasgow, F. & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, *70*, 662-680.

Dulaney, S. K., Hallam, P. R., & Wall, G. (2013). Superintendent perceptions of multi-tiered systems of support (MTSS): Obstacles and opportunities for school system reform. *AASA Jounral of Scholarship and Practice, 10*(2), 30-45.

Ernesto, M. (2013). Revised Standards for Educational and Psychological Testing to be published in spring 2014. *Psychological Science Agenda*, October 2013. Retrieved from, http://www.apa.org/science/about/psa/2013/revised-standards.aspx.

Fuchs, D. & Deschler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice, 22*, 129-136.

Fuchs, D. & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99.

Gaumer Erickson, A.S., Noonan, P.M., & Jenson, R. (2012). The School Implementation Scale: Measuring implementation in response to intervention models. *Learning Disabilities: A Contemporary Journal, 10*(2), 33-52.

Gliem, J.A. & Gliem, R.R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Green, S., & Salkind, N. J. (2011). *Spss quickstarts*. Pearson Higher Ed.

Gronlund, N.E., & Linn, R.L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*, *34*(3), 177.

Hooper, D., Coughlin, J., & Mullen, M.R. (2008). Structural equation modeling: Guidelines for determining model fit. *Journal of Business Research Methods, 6*(7), 53-60.

Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, *18*(3), 117-144.

Horner, R. H. (2000). Positive behavior supports. *Focus on autism and other disabilities, 15*(2), 97-105.

Horner, R. H., Sugai, G., Lewis-Palmer, T., & Todd, A. W. (2001). Teaching school-wide behavioral expectations. *Report on Emotional and Behavioral Disorders in Youth*, *1*(4), 77-79.

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Irvine, J.J. (1983). The accuracy of pre-service teachers' assessment of their classroom behaviors. *Journal of Research & Development in Education, 17,* 25-31.

Jenson, R. J. (2008). *Missouri integrated model implementation blueprint*. Kansas City, MO: University of Missouri-Kansas City, Institute for Human Development.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,*409-426.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices.

Kalberg, J.R., Lane, K.L., & Menzies, H.M. (2010). Using systematic screening procedures to identify students who are nonresponsive to primary prevention efforts: Integrating academic and behavioral measures. *Education and Treatment of Children, 33,* 561-584.

Kame'enui, E. J., & Simmons, D. C. (2000). Planning and evaluation tool for effective schoolwide reading programs. *Eugene, OR: Institute for the Development of Educational Achievement*.

Kimberlin, C.L. & Winterstein, A.G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy, 65*, 2276-2284.

Kincaid, D., Childs, K., Blasé, K. A., & Wallace, F. (2007). Identifying barriers and facilitators in implementing schoolwide positive behavior support. *Journal of Positive Behavior Interventions, 9*(3), 174-184.

Kirk, D. J. & Jones, T. L. (2004). Effective schools. *Pearson Assessment Report*. Pearson Education, Inc.

Lane, K.L., Kalberg, J.R., & Menzies, H.M. (2009). *Developing school-wide programs to prevent and manage problem behaviors: A step-by-step approach.* New York, NY: Guilford Press.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate behavioral research*, *32*(1), 53-76.

Liu, K., Alonzo, J., & Tindal, G. (2011). Implementation and outcomes. *Models for implementing response to intervention: Tools, outcomes, and implications*, 341-374.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). Statistical theories of mental test scores.

Mellard, D.F., Stern, A., & Woods, K. (2011). RTI school-based practices and evidence-based models. *Focus on Exceptional Children, 43*(6), 1-15.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525-543.

Messick. S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1990). Validity of test interpretation and use. In M. C. Alkin, (Ed.), Encyclopedia of Educational Research (6th ed.), New York: Macmillan, 1991.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist*, *50*(9), 741.

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research*, *26*(3), 479-497.

Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, *73*(3), 574.

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies*, *26*(3), 573-596.

Muthen, L.K. & Muthen, B.O. (2015). Mplus Statistical Modeling Program, Version 7.4. Los Angeles, CA.

Muthén, L.K. & Muthén, B.O. (1998-2015). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nunnally, J. (1978). *Psychometric theory*. McGraw-Hill, New York: NY.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6, 2*(11), 559-572.

Peter, J. P. & Churchill Jr, G. A. (1986). Relationships among research design choices and psychometric properties of rating scales: a meta-analysis. *Journal of Marketing Research*, 1-10.

Podsakoff, P. M., MacKenzie, S. B., & Podskoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, (65)*, 539-569.

Positive Behavioral Interventions and Supports website (https://www.pbis.org/school). What is school-wide positive behavioral interventions & supports? May 4, 2009.

The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users (3rd ed.). (2011). The Joint Committee on Standards for Education Evaluation. Yarbough, D.B., Shulla, L.M., Hopson, R.K., & Caruthers, F.A. (eds.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions, 15*(1), 39-50.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. Psychological Assessment, 12(3), 287-297.

Richardson, M. W. & Kuder, G. F. (1939). The calculation of test reliability coefficients based on the method of rational equivalence. *Journal of Educational Psychology, 30*(9), 681-687.

Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, *13*(4), 403-418.

Samuels, C. A. (2016). RTI's rapid (r)evolution. *Education Week, 36*(15), 4-9.

Sandomierski, T., Kincaid, D., & Algozzine, B. (2007). Response to intervention and positive behavior support: Brothers from different mothers or sisters with different misters? *Positive Behavioral Interventions and Supports Newsletter*, *4*(2), 1-4.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-337.

Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Berkeley: Houghton Mifflin Company.

Simonsen, B., Sugai, G., & Negron, M. (2008). Schoolwide positive behavior supports: Primary systems and practices. *Teaching Exceptional Children, 40*(6), 32-40.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies, 26,* 597–619.

Smith, G. T. & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7*, 300-308.

Sparks, S. (2016). Can this initiative be sustained? *Education Week, 36*, 15, 10-14.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25,* 78–90.

Steiger. J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research. 25,* l73- 180.

Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-898.

Sugai, G. & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality: A Special Education Journal, 17*(4), 223-237.

Sugai, G., Lewis-Palmer, T., Todd, A.W., & Horner, R.H. (1999). *School-Wide Evaluation Tool (SET).* Eugene, OR: University of Oregon.

Suzuki, S., & Rancer, A. S. (1994). Argumentativeness and verbal aggressiveness: Testing for conceptual and measurement equivalence across cultures. *Communication Monographs, 6,* 256–279.

Tabachnick, B. G. & Fidell, L. S. (2006). *Using Multivariate Statistics* (5th ed.). Boston: Allyn & Bacon.

Tabachnick, B. G. & Fidell, L.S. (2013). *Using Multivariate Statistics* (6th ed.). Boston: Pearson Education, Inc.

Todd, A. W., Lewis-Palmer, T., Horner, R. H., Sugai, G., Sampson, N. K., & Phillips, D. (2003). The school-wide evaluation tool: SET implementation manual. *University of Oregon*.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.

Walker, B.A. (2006). *The initial psychometric assessment of the BEACONS positive behavior support individual and team self-assessment and program review* (Unpublished doctoral dissertation). University of Washington, Seattle.

Welkenhuysen-Gybels, J., Billiet, J., & Cambre, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology, 34,* 702-722.

Widaman, K.F. & Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance abuse domain. *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*, 281-324.

Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Windle, M., Iwawaki, S., & Lerner, R.M. (1988). Cross-cultural comparability of temperament among Japanese and American preschool children. *International Journal of Psychology, 23,* 547–567.

Appendices

Appendix A: School Implementation Scale Instrument

School Implementation Scale

University of Kansas, Center for Research on Learning Patricia M. Noonan, PhD. | Amy Gaumer-Erickson, PhD.

Item	Scale: 5 (very true of me now) 4 3 (somewhat true of me now) 2 1 (not at all true of me now)
C1	I can summarize my school's shared vision/mission.
C2	I receive school-wide academic and behavioral data in usable and understandable formats.
C3	I have a clear understanding of the State Standards for my grade/subject.
C4	I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).
C5	I have the technology and resources that I need to provide effective instruction.
C6	I have the time necessary to analyze student data and problem solve with my colleagues.
C7	I am involved in action planning school-wide improvements with the other staff and administrators.
C8	I feel that my administrators are committed to implementing evidence-based instruction practices.
C9	I think that the current school initiatives are improving education for students in my school.
E1	My instruction intentionally addresses the State Standards for my grade/subject.
E2	I am able to differentiate instruction according to student needs while addressing the State Standards.
E3	I monitor each of my student's progress toward meeting the State Standards for my grade/subject.
E4	I review universal screening data at least three times a year for every student that I support.
E5	I review formative assessment data for every student that I support.
E6	I evaluate the effectiveness of my instruction based on assessment data.
E7	I adapt the environment, curriculum, and instruction based on each student's academic data.
E8	I adapt the environment, curriculum, and instruction based on each student's behavioral data.
E9	I modify my instructional practices based on students' formative assessment data.
E10	Based on assessment results, I re-teach information that students have not mastered.
E11	I am involved in meetings where data results are discussed.
E12	When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.
E13	When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.
F1	I consider my students' backgrounds when planning instruction.
F2	I regularly communicate with families regarding student academic goals/progress.
F3	I regularly communicate with families regarding student behavioral goals/progress.
F4	I make informed decisions based on feedback from families.
F5	I think my school does a good job of including parents as team members in data-based decision making.
P1	I participate in professional development where I learn how to develop curricular plans that address the State Standards.
P2	I participate in professional development where I learn strategies to improve my instructional practices.
P3	I participate in professional development where I learn how to monitor students' progress.
P4	I receive coaching/mentoring to implement evidence-based instructional practices.

C=School Culture, E=Evidence-Based Practices, F=Family Engagement, P=Professional Development

Appendix B: Accompanying Documentation for the School Implementation Scale

Measuring District-Wide Implementation of Academic and Behavior Supports with the *School Implementation Scale*

Amy Gaumer Erickson, Ph.D., University of Kansas
Pattie Noonan, Ph.D., University of Kansas

Gaumer Erickson, Noonan, & Jenson (2012) identified gaps in the existing fidelity measures being employed for integrated academic and behavior RTI models, stating that "there are no existing measures that evaluate the implementation of the core features of integrated models from a whole school perspective in a cost-effective minimally intrusive manner. Furthermore, measures that do exist are intervention specific and not appropriate for integrated models. Without treatment integrity data there is no way of understanding the variance in school gains" (p.36). They also claim that the data produced by existing treatment fidelity or implementation measures usually lacks utilization focus, meaning that the data are not visually represented in a meaningful and easily understandable way for continual planning and improvement purposes (Gaumer Erickson, Noonan & Jenson, 2012).

The *School Implementation Scale* was developed to encompass the evidence-based essential elements of effective school systems: 1) school culture, 2) ongoing professional development, 3) evidence-based practices, and 4) family engagement. These essential elements should be integrated into the school climate to drive decision making, facilitate innovation, and support students- The 33-item instrument has shown acceptable validity and reliability, and moderate correlations were found between SIS scores and reading/writing achievement of students with disabilities (Gaumer Erickson, Noonan & Jenson, 2012). The *School Implementation Scale* is an online survey designed to capture the personal adoption and level of individual implementation of tiered academic and behavior support. The survey is designed to be deployed school-wide (all instruction staff) as a way to measure levels of implementation over time.

The *School Implementation Scale*:

- Evaluates the fidelity and extent of implementation across a school and/or district.
- All items are evidence-based and result in robust data.
- Addresses tiered supports in both academics and behavior.
- Is sensitive to change.
- Is quick and easy to complete online. Results in a high response rate.
- Supports district- and school-level action planning and data-based decision-making.
- Correlates closely with gains in academic achievement for students with disabilities.
- Produces reliable results in elementary, middle, and high schools (overall scale Alpha = .961)
- Appeals to teachers, administrators, other certified staff, and noncertified staff.

When used in conjunction with other data, the *School Implementation Scale* yields powerful information about school improvement. Results can be utilization-focused to help schools/districts target specific areas for resources and professional development, making interventions more effective.

Example Triangulation of Data

- Process Checklists & Initiative-specific Fidelity Measures
- Staffing Changes
- Office Disciplinary Referrals
- Universal Screening/Progress Monitoring
- Proficiency on State Assessments
- Coaching Observations

Administering the *School Implementation Scale*

The *School Implementation Scale* is administered through a standardized process which includes providing instructions to each school's principal with sample text for email distribution to staff. To administer the survey, please email all your instructional staff asking them to complete this short online survey. Survey responses will be automatically graphed, with the report available on www.orspdgdata.org. All responses are confidential and will be aggregated in reporting. Staff accesses and completes the *School Implementation Scale* (33 items on a 5-point Likert-scale and 3 demographic items: school name, professional role, and membership on the school leadership team), which requires approximately 5-10 minutes. Below is sample text for the email to be sent to all instructional staff in your district. Please contact Dr. Amy Gaumer Erickson (aerickson@ku.edu) with questions regarding this survey.

> _____**(District/School Name)** Staff,
>
> As part of the _____**(name of program/model)** implementation process, it is important to get your input on the current status of implementation at our school. The School Implementation Scale will provide valuable data that will be used to improve education for all students in our district/school. Please go to http://www.orspdgdata.org/Surveys/ImplementationSurvey.php and complete this short survey. Submit your survey by _____**(one week from today**). The data will then be analyzed and shared with you on _____**(at the next inservice or via email)**.
>
> Thank you in advance for completing the survey.
> Sincerely,

Technical information regarding the *School Implementation Scale* can be found at:
Gaumer Erickson, A.S., Noonan, P.M., & Jenson, R. (2012). The *School Implementation Scale*: Measuring implementation in response to intervention models. *Learning Disabilities: A Contemporary Journal, 10*(2), 33-52.

School Implementation Scale Data Security

There are minimal potential risks or harm for the individual respondents, the various schools, districts, or states at which they are employed. Only the researchers will have access to any respondent-level information. It is possible, however, with Internet communications such as this questionnaire, that through intent or accident someone other than the intended recipient (the researcher) may see your response.

Individual respondent data will be confidential as the state, district, or schools will receive no individual level results; only overall reports and reports aggregated by the demographic questions at the beginning of the survey (school, staff role, and membership on the implementation or leadership team) will be disseminated. All data including demographic information and questionnaire responses and scores will be kept on a password secured computer in a locked office to which only the researchers will have access.

School Implementation Scale Reports and Interpretation

The *School Implementation Scale* Summary Reports provide aggregated data for your district and school regarding the essential elements of multi-tiered support systems. It is not expected that schools or districts will have high levels of implementation across all items. Instead, the results should be used to identify strengths and prioritize areas of improvement. Data reports from the *School Implementation Scale* have been utilized by both school and state teams for ongoing planning, refinement and improvement of their integrated academic and behavior RTI model; and the *School Implementation Scale* reports are meaningful to stakeholders and can be incorporated into data-based decision making.

School summary reports provide a table displaying score distributions and means for each item and domain. The number of respondents at each level is also added to the summary report, as well as the mode for each item is highlighted to provide visual representation for which response options rendered the highest percentage of staff. Mean scores by essential element domain, level of implementation sub-scales, and for each item are also presented as stacked bar graphs to increase the usefulness of the data.

The in-depth information on the school summary report can be used for action planning and to compare across years. Annual comparison reports were developed to illustrate the changes in the level of implementation subscales and of the essential elements domains and across school years. The annual comparison reports of the *School Implementation Scale* data related to the attainment of school-wide implementation is presented as 80% of staff or more with responses of 4 or 5 on each item as this data is more beneficial to schools for targeting areas for improvement.

Discussing the results of the *School Implementation Scale* within the implementation or leadership team and with the whole staff is encouraged to fully benefit from the information regarding the implementation status of the school. Identifying the strengths and weaknesses in the implementation of the integrated academic and behavior RTI model, can facilitate action plans for targeted improvement of the staff implementation of the model. Below are some examples of guided discussion questions to assist in the interpretation and processing of the *School Implementation Scale* results.

Guided Discussion Questions for District/School Leadership Teams

1. Quickly glance through the data. What are your first impressions?
2. Does the number/role of survey participants adequately represent our schools?
3. Celebrate successes: Which items or essential elements show high levels of implementation? What processes, professional development, etc. are in place that support these high levels of implementation?
4. How do the results from the *School Implementation Scale* align with other school-level data? Is additional data needed?
5. Prioritize needs: Which essential elements show low levels of implementation? Which survey items highlight areas that could be improved over the next year?
6. Next steps: How do the results influence our action planning for next year?

Ongoing Improvement of the *School Implementation Scale*

The data gathered through the administration of the *School Implementation Scale* is also used by the researchers for the continued improvement of the instrument in terms of additional validity and reliability evidence, and in terms of the relationship of the results with initial outcomes related to the implementation of integrated academic and behavior RTI models. The statistical analyses conducted with all *School Implementation Scale* data collected consists of reliability estimates in the form of Cronbach's alphas to calculate the internal consistency of the items, Confirmatory Factor Analysis to determine the factor structure of the items, Analyses of Variance (ANOVA) and T-tests to compare various group means or change over time, Pearson Product-Moment Correlations of *School Implementation Scale* results with student outcomes and other RTI model-specific measures of fidelity of implementation, and Hierarchical Regression Modeling to examine the relationships between *School Implementation Scale* results and individual-, school-, district-, and state-level characteristics or demographic factors.

Scale reliability is the proportion of variance attributable to the true score of the latent variables, and internal consistency reliability is concerned with the homogeneity of the items within the scale (DeVillis, 2003), or that the items are highly intercorrelated. This internal consistency is typically equated with Cronbach's (1951) coefficient alpha. Cronbach's coefficient alpha values are computed for the SIS overall as administered in each state, as well as for the four essential element domains and the three levels of implementation subscales.

Confirmatory factor analyses are conducted, first specifying three factors for the levels of implementation subscales then four factors for the essential elements domains. Principal Axis Factoring is used for the extraction of the factors because it uses an iterative process to estimate the communalities, and its goal is to extract the maximum orthogonal variance from the dataset with each succeeding factor (Tabachnick & Fidell, 2007). Rotation methods are categorized by whether they are orthogonal when the factors are not correlated, and oblique when the factors may be correlated. Assuming that there are correlations between the factors, Promax Rotation with Kaiser Normalization is employed for the rotation procedures because it rotates orthogonal factors into oblique positions (Tabachnick & Fidell, 2007) to ease interpretability of the factor structure. Item loadings on each factor for the three levels of implementation subscales and the four essential elements domains are interpreted according to the rules of thumb Comrey & Lee (1992) provided for interpreting factor loadings. They suggest that loadings higher than 0.71 are considered excellent, 0.63 very good, 0.55 good, 0.45 fair, and 0.32 poor.

The validity of the SIS is also examined by comparing the mean scores for various groups of participants or means for implementation level sub-scales and essential element domains. A One-Way Analysis of Variance (ANOVA) is conducted to compare means across the three levels of implementation (school, classroom, and individual student) and across the four essential elements domains. Significant differences detected by the ANOVAs are then subjected to follow-up Tukey tests to investigate where exactly the difference exists (Green & Salkind, 2011). These same procedures are conducted for differences between staff roles and participation on the implementation or leadership team.

Implementation across years is examined by conducting a series of Independent T-tests between the consecutive years' means for the full *School Implementation Scale*, the four essential elements domains, and the three levels of implementation. To further assess degree of implementation across years, participant ratings are converted to binomial scores where responses of 4 or 5 equal implementation and responses of 1, 2, and 3 equal lack of implementation. A threshold of 80% of school staff implementing each multi-tiered component was set as definition of school-wide implementation. This level is consistent with the research published by Sugai, et al. in 1999, and has also been reported by Simonsen, Sugai & Negron (2008) as the sufficient level to indicate school-wide implementation of the model.

The relationship between the School Implementation Scale results and behavioral/academic achievement for students with and without disabilities is investigated by computing the Pearson Product Moment Correlation. Using the degree of implementation method describe above, the percentage of school staff respondents that rated each item at 4 or 5 (threshold for school-wide implementation) is identified. The percent of the SIS items rated at level 4 or 5 by 80-100% of staff within each school is compared to the schools' change in percent of students that met proficiency on the state assessments or changes in other student outcomes, and correlations are computed. The *School Implementation Scale* results are also used to examine the relationship between the instrument results and any other specific fidelity measure results administered in the state, district, or school related to the particular RTI model being implemented.

To examine the relationship between individual respondent characteristics, school-, district-, and state-level characteristics with the School Implementation Scale results, hierarchical regression modeling is employed. A variety of variables including staff role, participation on implementation or leadership team, percentages of school students receiving free/reduced lunch or from ethnically diverse backgrounds, urbanicity classification of schools and districts, and geographical areas within the state are entered into a levelled regression equation. This renders a coefficient for each of the variables as to what extent they individually and jointly help predict results for the School Implementation Scale.

The final way in which the data will be utilized involves comparing the standardized measure results *School Implementation Scale* to compare across different integrated academic and behavior RTI models within and between states and/or districts.

The results of the research data analysis will be used for instrument improvement purposes. This research data is not associated with individuals, but linked to state, district and school numbers for matching and tracking purposes only. Any research published on the *School Implementation Scale* will be reported in aggregate anonymously by labels such as "State #1," and no state, district, or school names or numbers will be identified.

References:

- Comrey, A.L. & Lee, H.B. (1992). *A First Course in Factor Analysis* (2nd ed.). Hilldale, NJ: Lawrence Erlbaum Associates.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Devillis, Robert F. (2003). *Scale Development: Theory and Applications, 2nd edition*. Thousand Oaks, CA: Sage Publications.
- Gaumer Erickson, A.S., Noonan, P.M., & Jenson, R. (2012). The School Implementation Scale: Measuring implementation in response to intervention models. *Learning Disabilities: A Contemporary Journal, 10*(2), 33-52.
- Green, S.B. and Salkind, N.J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data, 6th edition*. Upper Saddle River, NJ: Pearson Education, Inc. publishing as Prentice Hall.
- Sugai, G., Lewis-Palmer, T., Todd, A.W., & Horner, R.H. (1999). *School-Wide Evaluation Tool (SET)*. Eugene, OR: University of Oregon.
- Simonsen, B., Sugai, G., & Negron, M. (2008). Schoolwide positive behavior supports: Primary systems and practices. *Teaching Exceptional Children, 40*(6), 32-40.
- Tabachnick, B.G. and Fidell, L.S. (2007). *Using Multivariate Statistics, 5th edition*. Boston, MA: Pearson Education, Inc. publishing as Allyn and Bacon.

Appendix C: Correlation Tables

Table 1. Midwestern State Collaborative Work 28 items: Inter-Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																											
C2	**.419**	1.000																										
C3	**.283**	**.275**	1.000																									
C4	**.521**	**.490**	**.260**	1.000																								
C5	**.284**	**.324**	**.195**	**.531**	1.000																							
C6	**.332**	**.529**	**.297**	**.452**	**.382**	1.000																						
C7	**.471**	**.509**	**.291**	**.501**	**.316**	**.406**	1.000																					
C8	**.373**	**.395**	**.178**	**.474**	**.273**	**.314**	**.377**	1.000																				
C9	**.419**	**.504**	**.276**	**.522**	**.402**	**.486**	**.479**	**.435**	1.000																			
E1	**.259**	**.272**	**.674**	**.223**	**.167**	**.229**	**.273**	**.197**	**.259**	1.000																		
E2	**.283**	**.333**	**.413**	**.330**	**.268**	**.350**	**.300**	**.302**	**.359**	**.386**	1.000																	
E3	**.309**	**.435**	**.482**	**.289**	**.203**	**.361**	**.361**	**.251**	**.357**	**.545**	**.433**	1.000																
E4	**.297**	**.484**	**.250**	**.264**	**.212**	**.374**	**.383**	**.194**	**.432**	**.266**	**.283**	**.460**	1.000															
E6	**.360**	**.537**	**.414**	**.329**	**.237**	**.413**	**.394**	**.273**	**.409**	**.476**	**.381**	**.579**	**.501**	1.000														
E7	**.270**	**.336**	**.364**	**.260**	**.191**	**.279**	**.288**	**.264**	**.327**	**.422**	**.509**	**.509**	**.366**	**.487**	1.000													
E8	**.283**	**.336**	**.317**	**.284**	**.195**	**.282**	**.302**	**.264**	**.316**	**.329**	**.481**	**.414**	**.328**	**.407**	**.614**	1.000												
E11	**.335**	**.621**	**.280**	**.385**	**.241**	**.460**	**.448**	**.366**	**.441**	**.307**	**.275**	**.406**	**.459**	**.543**	**.343**	**.285**	1.000											
E12	**.297**	**.396**	**.304**	**.308**	**.205**	**.368**	**.315**	**.250**	**.335**	**.346**	**.348**	**.415**	**.381**	**.461**	**.440**	**.376**	**.420**	1.000										
E13	**.314**	**.375**	**.264**	**.323**	**.199**	**.348**	**.327**	**.270**	**.331**	**.296**	**.325**	**.368**	**.326**	**.413**	**.425**	**.396**	**.397**	**.738**	1.000									
P1	**.322**	**.408**	**.420**	**.364**	**.239**	**.350**	**.416**	**.324**	**.376**	**.403**	**.341**	**.533**	**.343**	**.454**	**.369**	**.345**	**.393**	**.355**	**.303**	1.000								
P2	**.306**	**.373**	**.326**	**.366**	**.235**	**.302**	**.358**	**.415**	**.356**	**.342**	**.396**	**.401**	**.264**	**.383**	**.389**	**.388**	**.376**	**.365**	**.331**	**.555**	1.000							
P3	**.366**	**.503**	**.366**	**.423**	**.253**	**.411**	**.443**	**.371**	**.408**	**.346**	**.394**	**.554**	**.436**	**.468**	**.396**	**.403**	**.470**	**.419**	**.370**	**.653**	**.562**	1.000						
P4	**.362**	**.463**	**.284**	**.466**	**.281**	**.402**	**.441**	**.464**	**.436**	**.268**	**.353**	**.379**	**.317**	**.391**	**.322**	**.359**	**.410**	**.325**	**.304**	**.495**	**.501**	**.535**	1.000					
F1	**.324**	**.349**	**.324**	**.302**	**.285**	**.295**	**.340**	**.244**	**.373**	**.331**	**.374**	**.398**	**.382**	**.417**	**.452**	**.447**	**.298**	**.404**	**.383**	**.328**	**.323**	**.362**	**.306**	1.000				
F2	**.286**	**.350**	**.266**	**.231**	**.149**	**.268**	**.322**	**.184**	**.289**	**.277**	**.304**	**.433**	**.453**	**.462**	**.416**	**.378**	**.420**	**.365**	**.357**	**.297**	**.274**	**.342**	**.272**	**.381**	1.000			
F3	**.284**	**.358**	**.260**	**.237**	**.148**	**.253**	**.333**	**.174**	**.280**	**.284**	**.285**	**.432**	**.446**	**.460**	**.385**	**.396**	**.401**	**.355**	**.363**	**.282**	**.243**	**.320**	**.254**	**.382**	**.870**	1.000		
F4	**.336**	**.387**	**.248**	**.318**	**.217**	**.314**	**.345**	**.249**	**.363**	**.267**	**.321**	**.409**	**.433**	**.443**	**.413**	**.411**	**.413**	**.412**	**.423**	**.336**	**.302**	**.372**	**.323**	**.460**	**.641**	**.655**	1.000	
F5	**.426**	**.473**	**.201**	**.516**	**.365**	**.427**	**.430**	**.391**	**.470**	**.188**	**.316**	**.299**	**.363**	**.348**	**.301**	**.334**	**.377**	**.345**	**.353**	**.321**	**.299**	**.399**	**.411**	**.378**	**.385**	**.392**	**.538**	1.000

Correlation significant p<.01 (2-tailed)

Table 2. Western State EBISS 28 items: Inter-Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																											
C2	**.363**	1.000																										
C3	**.259**	**.247**	1.000																									
C4	**.386**	**.570**	**.154**	1.000																								
C5	**.207**	**.262**	**.150**	**.394**	1.000																							
C6	**.194**	**.326**	**.192**	**.311**	**.298**	1.000																						
C7	**.316**	**.441**	**.296**	**.335**	**.142**	**.282**	1.000																					
C8	**.394**	**.462**	**.241**	**.496**	**.235**	**.272**	**.474**	1.000																				
C9	**.369**	**.491**	**.301**	**.539**	**.271**	**.364**	**.457**	**.570**	1.000																			
E1	**.189**	**.199**	**.600**	**.089**	**.097**	**.141**	**.237**	**.175**	**.234**	1.000																		
E2	**.221**	**.254**	**.526**	**.178**	**.145**	**.265**	**.281**	**.247**	**.340**	**.716**	1.000																	
E3	**.268**	**.423**	**.370**	**.269**	**.148**	**.349**	**.496**	**.371**	**.454**	**.386**	**.466**	1.000																
E4	**.174**	**.389**	**.401**	**.171**	**.057**	**.222**	**.397**	**.267**	**.318**	**.471**	**.419**	**.521**	1.000															
E6	**.203**	**.302**	**.414**	**.171**	**.203**	**.205**	**.277**	**.253**	**.301**	**.449**	**.454**	**.385**	**.540**	1.000														
E7	**.207**	**.328**	**.371**	**.233**	**.122**	**.265**	**.313**	**.236**	**.312**	**.478**	**.573**	**.484**	**.542**	**.556**	1.000													
E8	**.243**	**.386**	**.339**	**.310**	**.177**	**.309**	**.345**	**.269**	**.370**	**.356**	**.469**	**.498**	**.473**	**.444**	**.748**	1.000												
E11	**.293**	**.624**	**.363**	**.389**	**.167**	**.317**	**.554**	**.438**	**.471**	**.326**	**.330**	**.483**	**.478**	**.349**	**.387**	**.410**	1.000											
E12	**.262**	**.382**	**.303**	**.324**	**.206**	**.387**	**.426**	**.336**	**.381**	**.313**	**.395**	**.538**	**.433**	**.394**	**.464**	**.453**	**.464**	1.000										
E13	**.291**	**.385**	**.252**	**.358**	**.218**	**.362**	**.399**	**.352**	**.379**	**.237**	**.325**	**.530**	**.330**	**.306**	**.386**	**.477**	**.428**	**.727**	1.000									
P1	**.242**	**.291**	**.518**	**.195**	**.187**	**.287**	**.261**	**.215**	**.310**	**.664**	**.597**	**.411**	**.423**	**.412**	**.443**	**.374**	**.381**	**.401**	**.327**	1.000								
P2	**.248**	**.358**	**.403**	**.260**	**.233**	**.242**	**.315**	**.292**	**.334**	**.467**	**.474**	**.375**	**.410**	**.445**	**.487**	**.440**	**.441**	**.452**	**.380**	**.543**	1.000							
P3	**.270**	**.446**	**.347**	**.337**	**.171**	**.426**	**.426**	**.352**	**.419**	**.381**	**.459**	**.532**	**.497**	**.424**	**.489**	**.467**	**.513**	**.526**	**.424**	**.568**	**.518**	1.000						
P4	**.253**	**.390**	**.265**	**.339**	**.228**	**.343**	**.335**	**.346**	**.378**	**.343**	**.385**	**.394**	**.400**	**.323**	**.388**	**.385**	**.421**	**.429**	**.383**	**.447**	**.512**	**.525**	1.000					
F1	**.142**	**.201**	**.376**	**.107**	**.081**	**.231**	**.265**	**.181**	**.238**	**.437**	**.525**	**.425**	**.374**	**.407**	**.504**	**.477**	**.315**	**.420**	**.377**	**.378**	**.423**	**.373**	**.251**	1.000				
F2	**.156**	**.231**	**.390**	**.096**	*.057*	**.155**	**.269**	**.171**	**.181**	**.475**	**.450**	**.398**	**.407**	**.402**	**.456**	**.367**	**.340**	**.337**	**.236**	**.411**	**.387**	**.399**	**.299**	**.428**	1.000			
F3	**.168**	**.295**	**.373**	**.117**	*.020*	**.209**	**.370**	**.229**	**.267**	**.426**	**.426**	**.525**	**.434**	**.350**	**.400**	**.441**	**.394**	**.397**	**.376**	**.407**	**.384**	**.429**	**.302**	**.462**	**.745**	1.000		
F4	**.157**	**.270**	**.382**	**.116**	*.039*	**.200**	**.363**	**.232**	**.276**	**.424**	**.451**	**.522**	**.437**	**.380**	**.443**	**.452**	**.358**	**.432**	**.398**	**.409**	**.405**	**.426**	**.265**	**.520**	**.668**	**.811**	1.000	
F5	**.290**	**.394**	**.183**	**.463**	**.260**	**.306**	**.252**	**.369**	**.413**	**.118**	**.213**	**.277**	**.136**	**.152**	**.219**	**.302**	**.304**	**.303**	**.358**	**.206**	**.242**	**.302**	**.323**	**.163**	**.151**	**.173**	**.194**	1.000

Correlation significant at p<.05 (2-tailed)

Correlation significant p<.01 (2-tailed)

Table 3. SW-PBS & SWPBIS 28 items: Inter-Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																											
C2	.246	1.000																										
C3	.273	.489	1.000																									
C4	.422	.420	.498	1.000																								
C5	.275	.546	.461	.469	1.000																							
C6	.234	.598	.488	.442	.750	1.000																						
C7	.186	.444	.319	.283	.354	.359	1.000																					
C8	.153	.629	.439	.370	.439	.521	.580	1.000																				
C9	.246	.468	.535	.481	.394	.406	.415	.530	1.000																			
E1	.354	.450	.328	.412	.517	.521	.394	.427	.405	1.000																		
E2	.344	.426	.484	.509	.435	.459	.316	.357	.588	.545	1.000																	
E3	.285	.468	.422	.435	.495	.539	.318	.424	.408	.515	.468	1.000																
E4	.348	.422	.307	.395	.490	.458	.311	.350	.342	.575	.413	.703	1.000															
E6	.322	.358	.228	.386	.333	.307	.267	.243	.323	.444	.444	.404	.421	1.000														
E7	.387	.293	.364	.411	.382	.385	.344	.315	.365	.470	.489	.453	.449	.426	1.000													
E8	.419	.323	.323	.449	.422	.386	.304	.277	.351	.463	.479	.415	.438	.445	.658	1.000												
E11	.248	.407	.397	.363	.442	.512	.374	.438	.416	.446	.455	.455	.383	.329	.454	.447	1.000											
E12	.364	.252	.212	.318	.307	.254	.243	.208	.262	.287	.330	.266	.298	.311	.316	.422	.327	1.000										
E13	.448	.291	.252	.428	.315	.258	.214	.170	.293	.301	.354	.263	.314	.368	.375	.502	.254	.436	1.000									
P1	.397	.286	.280	.376	.355	.317	.251	.234	.307	.451	.434	.392	.442	.352	.495	.486	.322	.418	.423	1.000								
P2	.198	.474	.377	.341	.476	.523	.327	.473	.420	.473	.403	.455	.412	.315	.394	.330	.434	.236	.205	.286	1.000							
P3	.227	.410	.313	.304	.474	.442	.322	.411	.361	.518	.377	.445	.461	.322	.399	.359	.378	.223	.215	.353	.770	1.000						
P4	.238	.403	.333	.292	.434	.445	.298	.377	.372	.490	.364	.441	.442	.300	.366	.329	.397	.226	.226	.340	.633	.725	1.000					
F1	.390	.296	.226	.386	.339	.284	.211	.193	.259	.379	.372	.326	.392	.407	.364	.437	.306	.378	.498	.332	.308	.315	.382	1.000				
F2	.246	.234	.229	.258	.220	.198	.169	.158	.204	.198	.235	.191	.196	.315	.204	.272	.250	.229	.453	.235	.138	.122	.152	.322	1.000			
F3	.169	.493	.387	.311	.479	.520	.325	.419	.358	.394	.349	.435	.357	.259	.286	.267	.397	.228	.188	.265	.443	.416	.466	.255	.215	1.000		
F4	.209	.311	.307	.341	.408	.450	.295	.327	.361	.488	.456	.441	.395	.322	.483	.461	.521	.270	.247	.387	.378	.411	.395	.290	.213	.356	1.000	
F5	.510	.398	.319	.450	.406	.359	.335	.289	.373	.472	.449	.382	.409	.463	.491	.557	.376	.426	.520	.474	.306	.324	.338	.455	.325	.295	.394	1.000

Correlation significant p<.01 (2-tailed)

Table 4. Response-to-Intervention 28 items: Inter-Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																											
C2	**.272**	1.000																										
C3	**.402**	**.403**	1.000																									
C4	**.437**	**.320**	**.595**	1.000																								
C5	**.411**	**.495**		**.471**	1.000																							
C6	**.302**	**.578**	**.468**	**.389**	**.716**	1.000																						
C7	**.244**	**.479**	**.357**	**.218**	**.409**	**.340**	1.000																					
C8	**.256**	**.695**	**.424**	**.352**	**.463**	**.507**	**.504**	1.000																				
C9	**.287**	**.587**	**.592**	**.446**	**.497**	**.434**	**.500**	**.618**	1.000																			
E1	**.405**	**.446**	**.358**			**.364**	**.309**	**.410**	*.489*	1.000																		
E2	**.370**	*.445*	**.535**		*.543*	**.462**	**.398**	**.487**	**.629**	**.535**	1.000																	
E3	**.392**	**.451**	**.377**	**.378**		**.379**	**.313**	**.406**	**.487**	**.602**	**.592**	1.000																
E4	**.411**	**.369**	**.408**	**.422**		**.359**	**.285**	**.345**	**.516**	**.572**	**.549**	**.668**	1.000															
E6	**.330**	**.354**	**.361**	**.356**	**.389**	**.346**	**.195**	**.262**	**.415**	**.407**	**.444**	**.395**	**.435**	1.000														
E7	**.519**	**.252**	**.480**	**.350**	*.379*	**.315**	**.443**	**.316**	**.386**	**.364**	**.420**	**.455**	**.436**	**.232**	1.000													
E8	**.389**	**.154**	**.381**	**.224**	**.294**	**.205**	**.409**	**.173**	**.303**	**.222**	**.326**	**.303**	**.302**	**.227**	**.633**	1.000												
E11	**.365**	**.391**	**.361**	**.246**	**.477**	**.407**	**.340**	**.383**	**.414**	**.412**	**.459**	**.411**	**.320**	**.295**	**.268**	**.267**	1.000											
E12	**.362**	**.312**	**.392**	**.282**	**.311**	**.205**	**.311**	**.291**	**.438**	**.303**	**.350**	**.268**	**.349**	**.261**	**.279**	**.269**	**.226**	1.000										
E13	**.486**	**.110**	**.264**	**.284**	**.324**	**.187**	**.345**	**.119**	**.207**	**.209**	**.280**	**.343**	**.367**	**.284**	**.495**	**.592**	**.212**	**.345**	1.000									
P1	**.558**	**.366**	**.427**	**.393**	**.431**	**.294**	**.371**	**.324**	**.418**	**.448**	**.436**	**.422**	**.434**	**.341**	**.545**	**.409**	**.335**	**.325**	**.394**	1.000								
P2	**.211**	**.432**	**.279**	**.273**	**.400**	**.409**	**.326**	**.495**	**.438**	**.482**	**.396**	**.357**	**.262**	**.160**	**.261**	**.133**	**.329**	**.142**	**.091**	**.229**	1.000							
P3	**.308**	**.455**	**.349**	**.319**	**.461**	**.401**	**.337**	**.492**	**.503**	**.546**	**.431**	**.425**	**.429**	**.276**	**.365**	**.161**	**.375**	**.234**	**.120**	**.348**	**.742**	1.000						
P4	**.223**	**.454**	**.348**	**.278**	**.467**	**.419**	**.359**	**.496**	**.467**	**.510**	**.457**	**.449**	**.418**	**.256**	**.251**	**.153**	**.385**	**.168**	**.077**	**.283**	**.646**	**.733**	1.000					
F1	**.407**	**.289**	**.337**	**.347**		**.219**	**.176**	**.128**	**.318**	**.344**	**.326**	**.300**	**.408**	**.382**	**.248**	**.321**	**.246**	**.351**	**.369**	**.321**	**.124**	**.192**	**.167**	1.000				
F2	*.273*	*.144*	**.280**			**.127**	**.109**	**.111**		**.119**	**.198**	**.208**	**.230**	**.371**	**.185**	**.242**	**.182**	**.192**	**.307**	**.213**	**.010**	**.066**	**.001**	**.252**	1.000			
F3	**.360**	**.521**	**.396**	*.342*		**.566**	**.416**	**.462**	**.456**	**.464**	**.469**	**.502**	**.419**	**.389**	**.304**	**.187**	**.417**	**.191**	**.205**	**.348**	**.392**	**.485**	**.516**	**.304**	**.104**	1.000		
F4	**.342**	*.434*	**.401**			**.447**	**.441**	**.532**	**.472**	**.509**	**.539**	**.536**	**.358**	**.240**	**.433**	**.287**	**.521**	**.215**	**.163**	**.385**	**.489**	**.498**	**.495**	**.107**	**-.032**	**.471**	1.000	
F5	**.551**	**.197**	**.303**	**.351**	**.417**	**.261**	**.353**	**.136**	**.183**	*.336*	**.291**	**.398**		**.330**	**.551**	**.557**	**.276**	**.340**	**.657**	**.482**	**.116**	**.213**	**.197**	**.390**	*.314*	**.314**	**.220**	1.000

Correlation significant at p<.05 (2-tailed)

Correlation significant p<.01 (2-tailed)

Table 5. Both States 28 items: Inter-Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																											
C2	.305	1.000																										
C3	.375	.462	1.000																									
C4	.451	.414	.527	1.000																								
C5	.313	.526	.474	.440	1.000																							
C6	.296	.590	.498	.424	.714	1.000																						
C7	.236	.475	.367	.308	.359	.389	1.000																					
C8	.222	.602	.448	.370	.423	.517	.608	1.000																				
C9	.313	.497	.567	.503	.416	.444	.464	.560	1.000																			
E1	.360	.486	.413	.435	.496	.527	.446	.502	.500	1.000																		
E2	.392	.457	.554	.549	.458	.468	.386	.414	.634	.569	1.000																	
E3	.323	.456	.449	.428	.477	.517	.347	.427	.434	.527	.503	1.000																
E4	.357	.406	.385	.398	.478	.464	.302	.353	.373	.524	.440	.736	1.000															
E6	.331	.361	.298	.409	.327	.319	.278	.244	.366	.399	.442	.399	.398	1.000														
E7	.411	.308	.393	.418	.344	.366	.354	.312	.386	.439	.487	.432	.410	.416	1.000													
E8	.425	.303	.352	.440	.353	.326	.294	.246	.364	.414	.470	.382	.375	.467	.631	1.000												
E11	.285	.410	.413	.390	.433	.504	.408	.453	.447	.499	.467	.448	.385	.362	.461	.454	1.000											
E12	.407	.286	.295	.353	.299	.274	.292	.249	.324	.320	.354	.311	.336	.312	.330	.411	.325	1.000										
E13	.496	.300	.324	.446	.327	.288	.273	.209	.329	.333	.402	.338	.365	.430	.418	.528	.304	.485	1.000									
P1	.434	.313	.347	.421	.337	.315	.310	.270	.375	.431	.454	.381	.380	.387	.496	.486	.356	.430	.462	1.000								
P2	.238	.433	.374	.344	.441	.486	.348	.440	.404	.497	.414	.438	.396	.280	.386	.304	.441	.275	.239	.315	1.000							
P3	.256	.405	.353	.332	.466	.448	.342	.413	.389	.530	.408	.444	.440	.288	.409	.339	.434	.282	.248	.362	.820	1.000						
P4	.274	.414	.374	.337	.449	.456	.318	.383	.400	.494	.418	.460	.445	.304	.389	.337	.419	.298	.275	.351	.662	.725	1.000					
F1	.409	.314	.297	.414	.358	.314	.233	.208	.316	.357	.396	.356	.397	.421	.375	.454	.314	.410	.515	.393	.342	.360	.427	1.000				
F2	.293	.251	.257	.303	.230	.215	.203	.182	.244	.241	.253	.240	.253	.378	.245	.319	.265	.283	.497	.269	.167	.158	.187	.359	1.000			
F3	.232	.482	.387	.325	.487	.506	.342	.421	.388	.414	.380	.453	.396	.302	.293	.291	.415	.270	.249	.295	.423	.432	.482	.303	.230	1.000		
F4	.216	.311	.305	.341	.365	.399	.290	.320	.358	.441	.442	.382	.305	.312	.447	.437	.508	.256	.230	.372	.397	.419	.411	.297	.172	.360	1.000	
F5	.500	.384	.363	.446	.378	.367	.351	.294	.378	.441	.434	.393	.397	.467	.483	.516	.384	.420	.553	.489	.307	.328	.354	.473	.380	.330	.375	1.000

Correlation significant p<.01 (2-tailed)

Table 6. 2014 Midwestern CW 31 Items- Inter Item Correlation Matrix

SIS Items	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.000																														
C2	.419	1.000																													
C3	.283	.275	1.000																												
C4	.521	.490	.260	1.000																											
C5	.284	.324	.195	.531	1.000																										
C6	.332	.529	.297	.452	.382	1.000																									
C7	.471	.509	.291	.501	.316	.406	1.000																								
C8	.373	.395	.178	.474	.273	.314	.377	1.000																							
C9	.419	.504	.276	.522	.402	.486	.479	.435	1.000																						
E1	.259	.272	.674	.223	.167	.229	.273	.197	.259	1.000																					
E2	.283	.333	.413	.330	.268	.350	.300	.302	.359	.386	1.000																				
E3	.309	.435	.482	.289	.203	.361	.361	.251	.357	.545	.433	1.000																			
E4	.297	.484	.250	.264	.212	.374	.383	.194	.432	.266	.283	.460	1.000																		
E5	.299	.437	.384	.262	.189	.332	.341	.254	.331	.455	.376	.637	.494	1.000																	
E6	.360	.537	.414	.329	.237	.413	.394	.273	.409	.476	.381	.579	.501	.638	1.000																
E7	.270	.336	.364	.260	.191	.279	.288	.264	.327	.422	.509	.509	.366	.514	.487	1.000															
E8	.283	.336	.317	.284	.195	.282	.302	.264	.316	.329	.481	.414	.328	.400	.407	.614	1.000														
E9	.297	.370	.390	.270	.194	.325	.310	.235	.350	.449	.436	.555	.435	.600	.635	.656	.500	1.000													
E10	.258	.318	.367	.229	.166	.286	.265	.230	.295	.414	.414	.517	.362	.541	.517	.566	.445	.617	1.000												
E11	.335	.621	.280	.385	.241	.460	.448	.366	.441	.307	.275	.406	.459	.467	.543	.343	.285	.389	.317	1.000											
E12	.297	.396	.304	.308	.205	.368	.315	.250	.335	.346	.348	.415	.381	.463	.461	.440	.376	.417	.414	.420	1.000										
E13	.314	.375	.264	.323	.199	.348	.327	.270	.331	.296	.325	.368	.326	.372	.413	.425	.396	.388	.369	.397	.738	1.000									
P1	.322	.408	.420	.364	.239	.350	.416	.324	.376	.403	.341	.533	.343	.451	.454	.369	.345	.404	.350	.393	.355	.303	1.000								
P2	.306	.373	.326	.366	.235	.302	.358	.415	.356	.342	.396	.401	.264	.396	.383	.389	.388	.389	.355	.376	.365	.331	.555	1.000							
P3	.366	.503	.366	.423	.253	.411	.443	.371	.408	.346	.394	.554	.436	.529	.468	.396	.403	.448	.404	.470	.419	.370	.653	.562	1.000						
P4	.362	.463	.284	.466	.281	.402	.441	.464	.436	.268	.353	.379	.317	.357	.391	.322	.359	.373	.303	.410	.325	.304	.495	.501	.535	1.000					
F1	.324	.349	.324	.302	.285	.295	.340	.244	.373	.331	.374	.398	.382	.380	.417	.452	.447	.426	.381	.298	.404	.383	.328	.323	.362	.306	1.000				
F2	.286	.350	.266	.231	.149	.268	.322	.184	.289	.277	.304	.433	.453	.461	.462	.416	.378	.414	.363	.420	.365	.357	.297	.274	.342	.272	.381	1.000			
F3	.284	.358	.260	.237	.148	.253	.333	.174	.280	.284	.285	.432	.446	.437	.460	.385	.396	.398	.355	.401	.355	.363	.282	.243	.320	.254	.382	.870	1.000		
F4	.336	.387	.248	.318	.217	.314	.345	.249	.363	.267	.321	.409	.433	.408	.443	.413	.411	.407	.346	.413	.412	.423	.336	.302	.372	.323	.460	.641	.655	1.000	
F5	.426	.473	.201	.516	.365	.427	.430	.391	.470	.188	.316	.299	.363	.275	.348	.301	.334	.320	.269	.377	.345	.353	.321	.299	.399	.411	.378	.385	.392	.538	1.000

Correlation significant p<.01 (2-tailed)

Appendix D: Mplus Confirmatory Factor Analysis Tables

Table 1. Midwestern CW 28 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Factor #1	C1	0.572	0.016	36.276	0.000
	C2	0.829	0.018	45.285	0.000
	C3	0.338	0.013	25.144	0.000
	C4	0.722	0.016	44.566	0.000
	C5	0.569	0.020	29.112	0.000
	C6	0.807	0.021	38.708	0.000
	C7	0.759	0.018	41.597	0.000
	C8	0.430	0.013	33.056	0.000
	C9	0.717	0.016	43.572	0.000
Factor #2	E1	0.374	0.011	33.279	0.000
	E2	0.447	0.013	34.273	0.000
	E3	0.606	0.013	45.447	0.000
	E4	0.797	0.022	36.467	0.000
	E6	0.668	0.014	47.443	0.000
	E7	0.457	0.011	41.516	0.000
	E8	0.471	0.013	36.880	0.000
	E11	0.656	0.017	38.384	0.000
	E12	0.494	0.013	39.520	0.000
	E13	0.438	0.012	36.768	0.000
Factor #3	P1	0.754	0.016	48.035	0.000
	P2	0.514	0.012	42.779	0.000
	P3	0.809	0.015	52.855	0.000
	P4	0.733	0.018	41.205	0.000
Factor #4	F1	0.364	0.014	26.546	0.000
	F2	0.824	0.013	65.853	0.000
	F3	0.870	0.013	66.391	0.000
	F4	0.670	0.014	46.254	0.000
	F5	0.541	0.020	27.075	0.000
Factor #3	Factor #1	0.779	0.010	76.848	0.000
Factor #3	Factor #1	0.779	0.011	73.131	0.000
Factor #3	Factor #2	0.794	0.010	80.330	0.000
Factor #4	Factor #1	0.505	0.016	32.112	0.000
Factor #4	Factor #2	0.689	0.012	59.255	0.000
Factor #4	Factor #3	0.463	0.017	27.836	0.000

Table 2. Midwestern CW 28 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.066	0.006	0.016	0.001
C2	0.085	0.008	0.021	0.001
C3	0.047	0.004	0.012	0.000
C4	0.094	0.009	0.023	0.001
C5	0.039	0.004	0.010	0.000
C6	0.056	0.005	0.014	0.000
C7	0.072	0.007	0.018	0.001
C8	0.070	0.006	0.017	0.001
C9	0.087	0.008	0.021	0.001
E1	0.011	0.042	0.021	0.004
E2	0.010	0.038	0.019	0.004
E3	0.016	0.060	0.030	0.006
E4	0.007	0.025	0.012	0.003
E6	0.016	0.062	0.031	0.007
E7	0.016	0.062	0.031	0.007
E8	0.011	0.043	0.022	0.005
E11	0.009	0.035	0.017	0.004
E12	0.013	0.050	0.025	0.005
E13	0.012	0.047	0.023	0.005
P1	0.022	0.015	0.174	-0.001
P2	0.022	0.015	0.177	-0.001
P3	0.029	0.020	0.232	-0.001
P4	0.014	0.010	0.112	-0.001
F1	0.001	0.003	-0.001	0.019
F2	0.005	0.023	-0.006	0.162
F3	0.005	0.023	-0.006	0.164
F4	0.001	0.006	-0.002	0.043
F5	0.000	0.002	-0.001	0.014

Table 3. Midwestern CW 28 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	.865																											
C2	.474	1.294																										
C3	.193	.280	.558																									
C4	.413	.599	.244	.999																								
C5	.326	.472	.192	.411	1.226																							
C6	.462	.669	.273	.583	.459	1.553																						
C7	.435	.630	.257	.549	.432	.612	1.234																					
C8	.246	.356	.145	.311	.245	.347	.327	.568																				
C9	.410	.594	.242	.518	.408	.578	.545	.308	1.026																			
E1	.167	.242	.099	.211	.166	.235	.221	.125	.209	.433																		
E2	.199	.289	.118	.251	.198	.281	.264	.150	.250	.167	.588																	
E3	.270	.392	.160	.341	.269	.381	.359	.203	.339	.227	.271	.699																
E4	.355	.514	.210	.448	.353	.501	.471	.267	.445	.298	.356	.483	1.692															
E6	.298	.431	.176	.376	.296	.420	.395	.224	.373	.250	.298	.405	.532	.799														
E7	.204	.295	.120	.257	.202	.287	.270	.153	.255	.171	.204	.277	.364	.305	.451													
E8	.210	.304	.124	.265	.209	.296	.279	.158	.263	.176	.211	.286	.375	.315	.215	.580												
E11	.293	.424	.173	.369	.291	.413	.388	.220	.367	.246	.293	.398	.523	.438	.300	.309	1.054											
E12	.220	.319	.130	.278	.219	.311	.292	.166	.276	.185	.221	.300	.394	.330	.226	.233	.325	.571										
E13	.195	.283	.115	.247	.194	.276	.259	.147	.245	.164	.196	.266	.349	.293	.200	.207	.288	.217	.502									
P1	.336	.487	.199	.424	.334	.474	.446	.253	.421	.224	.267	.363	.477	.400	.273	.282	.393	.296	.262	.970								
P2	.229	.332	.135	.289	.228	.323	.304	.172	.287	.153	.182	.247	.325	.272	.186	.192	.268	.202	.179	.387	.533							
P3	.361	.522	.213	.455	.359	.508	.479	.271	.452	.240	.287	.389	.511	.429	.293	.303	.422	.317	.282	.610	.415	.978						
P4	.327	.473	.193	.412	.325	.460	.433	.245	.409	.218	.260	.353	.463	.388	.266	.274	.382	.288	.255	.552	.376	.593	1.143					
F1	.105	.152	.062	.133	.104	.148	.139	.079	.132	.094	.112	.152	.200	.167	.114	.118	.164	.124	.110	.127	.086	.136	.123	.605				
F2	.238	.345	.141	.301	.237	.336	.316	.179	.298	.213	.254	.344	.452	.379	.259	.268	.373	.281	.249	.288	.196	.309	.280	.300	.807			
F3	.251	.364	.149	.317	.250	.354	.334	.189	.315	.225	.268	.364	.478	.401	.274	.283	.394	.296	.263	.304	.207	.326	.295	.316	.717	.890		
F4	.194	.281	.114	.244	.193	.273	.257	.146	.243	.173	.206	.280	.368	.308	.211	.218	.303	.228	.203	.234	.159	.251	.227	.244	.552	.583	.838	
F5	.156	.226	.092	.197	.155	.220	.207	.117	.196	.139	.167	.226	.297	.249	.170	.176	.245	.184	.163	.189	.129	.202	.183	.197	.446	.471	.362	1.288

Table 4. Western EBISS 28 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
Factor #1	C1	0.548	0.027	20.015	0.000
	C2	0.823	0.029	28.445	0.000
	C3	0.430	0.028	15.470	0.000
	C4	0.703	0.026	27.159	0.000
	C5	0.486	0.034	14.310	0.000
	C6	0.625	0.035	17.951	0.000
	C7	0.828	0.034	24.057	0.000
	C8	0.700	0.025	27.896	0.000
	C9	0.804	0.026	30.932	0.000
Factor #2	E1	0.721	0.029	24.637	0.000
	E2	0.760	0.027	28.104	0.000
	E3	0.900	0.030	30.078	0.000
	E4	0.947	0.034	27.682	0.000
	E6	0.629	0.025	25.635	0.000
	E7	0.745	0.023	31.761	0.000
	E8	0.737	0.024	30.344	0.000
	E11	0.785	0.031	25.504	0.000
	E12	0.786	0.028	28.563	0.000
	E13	0.704	0.028	24.861	0.000
Factor #3	P1	0.935	0.032	28.900	0.000
	P2	0.754	0.026	28.866	0.000
	P3	0.976	0.030	32.401	0.000
	P4	0.853	0.033	26.026	0.000
Factor #4	F1	0.562	0.025	22.274	0.000
	F2	0.990	0.029	34.480	0.000
	F3	1.141	0.027	41.595	0.000
	F4	1.073	0.027	39.741	0.000
	F5	0.280	0.032	8.784	0.000
Factor #2	Factor #1	0.723	0.018	40.874	0.000
Factor #3	Factor #1	0.702	0.020	35.034	0.000
Factor #3	Factor #2	0.891	0.011	79.224	0.000
Factor #4	Factor #1	0.437	0.026	16.783	0.000
Factor #4	Factor #2	0.736	0.016	46.362	0.000
Factor #4	Factor #3	0.634	0.021	30.102	0.000

Table 5. Western EBISS 28 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.050	0.007	0.009	-0.001
C2	0.084	0.013	0.016	-0.001
C3	0.036	0.005	0.007	-0.001
C4	0.088	0.013	0.017	-0.002
C5	0.027	0.004	0.005	0.000
C6	0.035	0.005	0.007	-0.001
C7	0.053	0.008	0.010	-0.001
C8	0.094	0.014	0.018	-0.002
C9	0.113	0.017	0.021	-0.002
E1	0.007	0.047	0.027	0.006
E2	0.009	0.063	0.036	0.008
E3	0.010	0.064	0.037	0.008
E4	0.007	0.048	0.028	0.006
E6	0.009	0.058	0.034	0.007
E7	0.014	0.094	0.054	0.012
E8	0.012	0.081	0.047	0.010
E11	0.007	0.046	0.027	0.006
E12	0.009	0.063	0.037	0.008
E13	0.007	0.048	0.028	0.006
P1	0.009	0.027	0.130	0.002
P2	0.011	0.033	0.160	0.003
P3	0.012	0.037	0.177	0.003
P4	0.007	0.021	0.103	0.002
F1	-0.001	0.008	0.003	0.042
F2	-0.002	0.015	0.006	0.084
F3	-0.005	0.031	0.014	0.178
F4	-0.004	0.026	0.011	0.146
F5	0.000	0.002	0.001	0.011

Table 6. Western EBISS 28 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.055																											
C2	.451	1.349																										
C3	.236	.354	.996																									
C4	.385	.579	.302	1.044																								
C5	.266	.400	.209	.342	1.466																							
C6	.342	.514	.269	.439	.303	1.632																						
C7	.454	.681	.356	.582	.402	.517	1.757																					
C8	.383	.576	.301	.492	.340	.437	.579	1.002																				
C9	.441	.662	.346	.566	.391	.502	.666	.563	1.137																			
E1	.285	.429	.224	.366	.253	.325	.431	.364	.419	1.318																		
E2	.301	.452	.236	.386	.267	.343	.455	.384	.442	.548	1.206																	
E3	.356	.535	.280	.457	.316	.406	.538	.455	.523	.649	.684	1.542																
E4	.375	.563	.294	.481	.333	.428	.567	.479	.551	.683	.720	.853	1.921															
E6	.249	.374	.196	.320	.221	.284	.377	.318	.366	.454	.478	.567	.596	.956														
E7	.295	.443	.232	.379	.262	.336	.446	.377	.433	.537	.567	.671	.706	.469	.967													
E8	.292	.438	.229	.374	.259	.332	.441	.372	.428	.531	.560	.663	.698	.464	.549	1.012												
E11	.311	.467	.244	.399	.276	.354	.470	.397	.456	.566	.597	.707	.744	.494	.585	.578	1.504											
E12	.311	.468	.244	.400	.276	.355	.470	.397	.457	.567	.598	.708	.745	.495	.586	.579	.617	1.265										
E13	.279	.418	.219	.358	.247	.318	.421	.356	.409	.507	.535	.633	.667	.443	.525	.518	.552	.553	1.260									
P1	.360	.540	.282	.462	.319	.410	.544	.459	.528	.601	.634	.751	.790	.525	.622	.614	.655	.656	.587	1.677								
P2	.290	.435	.228	.372	.257	.331	.438	.370	.426	.484	.511	.605	.637	.423	.501	.495	.528	.528	.473	.705	1.095							
P3	.375	.564	.295	.482	.333	.428	.567	.479	.551	.627	.661	.783	.824	.548	.649	.641	.683	.684	.612	.913	.736	1.569						
P4	.328	.493	.257	.421	.291	.374	.496	.419	.481	.548	.578	.684	.720	.478	.567	.560	.597	.598	.535	.798	.643	.832	1.650					
F1	.135	.202	.106	.173	.119	.153	.203	.172	.197	.298	.314	.372	.392	.260	.308	.305	.325	.325	.291	.333	.268	.347	.304	.928				
F2	.237	.356	.186	.304	.210	.270	.358	.303	.348	.525	.554	.656	.691	.459	.543	.537	.572	.573	.513	.587	.473	.613	.535	.556	1.525			
F3	.273	.411	.215	.351	.242	.312	.413	.349	.401	.606	.638	.756	.796	.529	.626	.619	.660	.661	.591	.677	.545	.706	.617	.641	1.130	1.600		
F4	.257	.386	.202	.330	.228	.293	.388	.328	.377	.569	.600	.711	.748	.497	.589	.582	.620	.621	.556	.636	.513	.664	.580	.603	1.063	1.225	1.491	
F5	.067	.101	.053	.086	.059	.076	.101	.086	.098	.148	.156	.185	.195	.130	.153	.152	.162	.162	.145	.166	.134	.173	.151	.157	.277	.319	.300	1.288

Table 7. SW-PBS/SWPBIS 28 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Factor #1	C1	0.541	0.023	23.632	0.000
	C2	0.813	0.025	32.694	0.000
	C3	0.374	0.022	17.382	0.000
	C4	0.684	0.024	28.432	0.000
	C5	0.471	0.028	16.607	0.000
	C6	0.757	0.030	24.975	0.000
	C7	0.786	0.028	28.184	0.000
	C8	0.531	0.021	25.573	0.000
	C9	0.802	0.024	34.137	0.000
Factor #2	E1	0.522	0.021	25.342	0.000
	E2	0.591	0.021	28.517	0.000
	E3	0.844	0.024	35.061	0.000
	E4	0.794	0.029	26.928	0.000
	E6	0.576	0.020	28.170	0.000
	E7	0.612	0.018	33.351	0.000
	E8	0.650	0.020	31.968	0.000
	E11	0.729	0.025	29.070	0.000
	E12	0.670	0.021	32.029	0.000
	E13	0.653	0.021	30.812	0.000
Factor #3	P1	0.785	0.025	30.971	0.000
	P2	0.571	0.020	28.840	0.000
	P3	0.891	0.025	35.505	0.000
	P4	0.821	0.028	29.522	0.000
Factor #4	F1	0.429	0.021	20.558	0.000
	F2	0.866	0.022	39.993	0.000
	F3	0.941	0.022	43.421	0.000
	F4	0.820	0.022	37.796	0.000
	F5	0.500	0.029	17.007	0.000
Factor #2	Factor #1	0.792	0.014	58.519	0.000
Factor #3	Factor #1	0.768	0.016	48.065	0.000
Factor #3	Factor #2	0.832	0.013	65.814	0.000
Factor #4	Factor #1	0.556	0.021	26.022	0.000
Factor #4	Factor #2	0.764	0.014	55.076	0.000
Factor #4	Factor #3	0.595	0.021	28.437	0.000

Table 8. SW-PBS/SWPBIS 28 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.055	0.007	0.013	0.000
C2	0.089	0.012	0.022	0.000
C3	0.040	0.005	0.010	0.000
C4	0.072	0.010	0.017	0.000
C5	0.029	0.004	0.007	0.000
C6	0.045	0.006	0.011	0.000
C7	0.060	0.008	0.015	0.000
C8	0.069	0.009	0.017	0.000
C9	0.105	0.014	0.026	0.000
E1	0.009	0.041	0.019	0.006
E2	0.011	0.049	0.023	0.008
E3	0.014	0.063	0.030	0.010
E4	0.007	0.032	0.015	0.005
E6	0.011	0.049	0.023	0.008
E7	0.017	0.075	0.035	0.012
E8	0.014	0.062	0.029	0.010
E11	0.009	0.042	0.020	0.007
E12	0.013	0.060	0.028	0.010
E13	0.012	0.055	0.026	0.009
P1	0.014	0.016	0.142	0.001
P2	0.016	0.018	0.157	0.001
P3	0.019	0.022	0.193	0.001
P4	0.011	0.013	0.116	0.001
F1	0.000	0.006	0.001	0.033
F2	0.000	0.021	0.003	0.112
F3	0.000	0.029	0.005	0.157
F4	0.000	0.017	0.003	0.093
F5	0.000	0.003	0.001	0.018

Table 9. SW-PBS & SWPBIS 28 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	.888																											
C2	.440	1.214																										
C3	.202	.304	.713																									
C4	.370	.556	.256	1.046																								
C5	.255	.383	.176	.322	1.217																							
C6	.409	.615	.283	.517	.356	1.583																						
C7	.425	.639	.294	.537	.370	.595	1.414																					
C8	.287	.431	.198	.363	.250	.401	.417	.748																				
C9	.434	.653	.300	.549	.378	.607	.631	.426	1.107																			
E1	.224	.336	.155	.283	.195	.313	.325	.220	.332	.759																		
E2	.253	.381	.175	.320	.221	.354	.368	.248	.376	.309	.808																	
E3	.362	.543	.250	.457	.315	.506	.525	.355	.536	.441	.499	1.222																
E4	.340	.511	.235	.430	.296	.476	.494	.333	.504	.415	.469	.670	1.583															
E6	.247	.371	.171	.312	.215	.345	.359	.242	.366	.301	.341	.486	.457	.779														
E7	.262	.394	.181	.332	.228	.367	.381	.257	.389	.320	.362	.517	.486	.353	.685													
E8	.279	.419	.193	.352	.243	.390	.405	.273	.413	.340	.384	.549	.516	.375	.398	.824												
E11	.312	.469	.216	.395	.272	.437	.454	.306	.463	.381	.431	.615	.578	.420	.446	.474	1.191											
E12	.287	.431	.198	.363	.250	.401	.417	.281	.426	.350	.396	.565	.531	.386	.410	.436	.488	.869										
E13	.280	.421	.193	.354	.244	.391	.406	.274	.415	.341	.386	.551	.518	.376	.400	.425	.476	.437	.874									
P1	.326	.490	.226	.412	.284	.456	.474	.320	.484	.341	.386	.551	.518	.376	.400	.425	.476	.437	.426	1.185								
P2	.237	.357	.164	.300	.207	.332	.345	.233	.352	.248	.281	.401	.377	.274	.291	.309	.346	.318	.310	.448	.699							
P3	.370	.557	.256	.468	.322	.518	.538	.363	.549	.387	.438	.625	.588	.427	.454	.482	.540	.496	.484	.699	.509	1.267						
P4	.341	.513	.236	.431	.297	.477	.496	.335	.506	.357	.404	.576	.542	.394	.418	.444	.498	.457	.446	.644	.469	.732	1.404					
F1	.129	.194	.089	.163	.112	.180	.187	.126	.191	.171	.194	.277	.260	.189	.201	.213	.239	.219	.214	.200	.146	.227	.210	.695				
F2	.260	.391	.180	.329	.227	.364	.378	.255	.386	.346	.391	.559	.525	.381	.405	.431	.482	.443	.432	.405	.294	.459	.423	.371	1.056			
F3	.283	.425	.196	.358	.246	.396	.411	.277	.420	.376	.425	.607	.571	.414	.440	.468	.524	.482	.470	.440	.320	.499	.460	.403	.815	1.123		
F4	.247	.371	.171	.312	.215	.345	.358	.242	.366	.328	.371	.529	.498	.361	.384	.408	.457	.420	.409	.383	.279	.435	.401	.352	.710	.772	1.022	
F5	.150	.226	.104	.190	.131	.210	.218	.147	.223	.200	.226	.322	.303	.220	.234	.248	.278	.256	.249	.233	.170	.265	.244	.214	.433	.470	.410	1.340

Table 10. RTI 28 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Two-Tailed Est./S.E.	P-Value
Factor #1	C1	0.545	0.065	8.336	0.000
	C2	0.625	0.056	11.142	0.000
	C3	0.487	0.062	7.867	0.000
	C4	0.715	0.059	12.212	0.000
	C5	0.514	0.077	6.712	0.000
	C6	0.730	0.079	9.200	0.000
	C7	0.837	0.074	11.307	0.000
	C8	0.733	0.057	12.947	0.000
	C9	0.849	0.060	14.187	0.000
Factor #2	E1	0.843	0.066	12.833	0.000
	E2	0.783	0.064	12.286	0.000
	E3	0.871	0.063	13.796	0.000
	E4	0.906	0.072	12.502	0.000
	E6	0.577	0.057	10.180	0.000
	E7	0.758	0.055	13.811	0.000
	E8	0.735	0.053	13.815	0.000
	E11	0.774	0.067	11.631	0.000
	E12	0.806	0.061	13.129	0.000
	E13	0.742	0.059	12.635	0.000
Factor #3	P1	1.031	0.073	14.046	0.000
	P2	0.798	0.062	12.783	0.000
	P3	1.006	0.068	14.777	0.000
	P4	0.777	0.073	10.584	0.000
Factor #4	F1	0.603	0.057	10.515	0.000
	F2	1.043	0.064	16.367	0.000
	F3	1.136	0.062	18.262	0.000
	F4	1.056	0.061	17.372	0.000
	F5	0.296	0.072	4.121	0.000
Factor #2	Factor #1	0.753	0.037	20.390	0.000
Factor #3	Factor #1	0.713	0.044	16.083	0.000
Factor #3	Factor #2	0.902	0.023	38.505	0.000
Factor #4	Factor #1	0.422	0.060	7.050	0.000
Factor #4	Factor #2	0.782	0.032	24.722	0.000
Factor #4	Factor #3	0.654	0.046	14.338	0.000

Table 11. RTI 28 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.040	0.009	0.007	-0.002
C2	0.072	0.017	0.013	-0.004
C3	0.042	0.010	0.008	-0.002
C4	0.083	0.019	0.015	-0.004
C5	0.026	0.006	0.005	-0.001
C6	0.038	0.009	0.007	-0.002
C7	0.056	0.013	0.010	-0.003
C8	0.094	0.022	0.017	-0.005
C9	0.116	0.027	0.021	-0.006
E1	0.012	0.062	0.038	0.011
E2	0.011	0.059	0.036	0.010
E3	0.014	0.074	0.046	0.013
E4	0.010	0.053	0.032	0.009
E6	0.009	0.047	0.029	0.008
E7	0.016	0.087	0.053	0.015
E8	0.016	0.087	0.053	0.015
E11	0.009	0.050	0.031	0.009
E12	0.013	0.068	0.042	0.012
E13	0.012	0.064	0.039	0.011
P1	0.009	0.037	0.167	0.002
P2	0.009	0.035	0.158	0.002
P3	0.011	0.044	0.195	0.003
P4	0.005	0.020	0.091	0.001
F1	-0.003	0.012	0.003	0.047
F2	-0.007	0.028	0.006	0.110
F3	-0.011	0.045	0.009	0.174
F4	-0.009	0.036	0.008	0.140
F5	-0.001	0.003	0.001	0.012

Table 12. RTI 28 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	1.165																											
C2	.341	.937																										
C3	.265	.304	.979																									
C4	.390	.446	.348	1.055																								
C5	.280	.321	.250	.367	1.510																							
C6	.398	.456	.355	.522	.375	1.754																						
C7	.457	.523	.407	.599	.430	.611	1.655																					
C8	.400	.458	.357	.524	.377	.535	.614	1.033																				
C9	.463	.530	.413	.607	.436	.620	.711	.622	1.186																			
E1	.346	.396	.309	.454	.326	.463	.532	.466	.539	1.399																		
E2	.321	.368	.287	.421	.303	.430	.493	.432	.500	.660	1.292																	
E3	.357	.409	.319	.469	.337	.479	.549	.481	.556	.734	.682	1.352																
E4	.372	.426	.332	.487	.350	.498	.571	.500	.579	.764	.709	.789	1.695															
E6	.237	.271	.211	.310	.223	.317	.363	.318	.368	.486	.451	.502	.522	.952														
E7	.311	.356	.278	.408	.293	.416	.478	.418	.484	.639	.593	.660	.686	.437	1.017													
E8	.302	.346	.269	.396	.284	.404	.463	.406	.470	.620	.575	.640	.666	.424	.557	.969												
E11	.317	.364	.283	.416	.299	.425	.488	.427	.494	.652	.605	.674	.701	.446	.586	.569	1.385											
E12	.331	.379	.295	.434	.312	.443	.508	.445	.515	.680	.631	.702	.730	.465	.611	.593	.624	1.250										
E13	.304	.349	.272	.399	.287	.407	.468	.409	.474	.625	.580	.646	.672	.428	.562	.545	.574	.598	1.137									
P1	.401	.460	.358	.526	.378	.537	.616	.540	.624	.784	.728	.810	.842	.536	.705	.684	.719	.749	.690	1.769								
P2	.310	.356	.277	.407	.292	.415	.477	.417	.483	.607	.563	.627	.652	.415	.545	.529	.557	.580	.534	.823	1.214							
P3	.391	.448	.349	.513	.369	.524	.601	.526	.609	.765	.710	.790	.821	.523	.687	.667	.702	.731	.673	1.038	.803	1.599						
P4	.302	.346	.270	.396	.285	.405	.464	.407	.471	.591	.548	.610	.635	.404	.531	.515	.542	.565	.520	.802	.620	.782	1.575					
F1	.139	.159	.124	.182	.131	.186	.213	.187	.216	.398	.369	.411	.427	.272	.357	.347	.365	.380	.350	.407	.315	.397	.307	.966				
F2	.240	.275	.214	.315	.226	.321	.369	.323	.374	.688	.638	.710	.738	.470	.618	.600	.631	.657	.605	.703	.544	.686	.530	.629	1.534			
F3	.261	.299	.233	.342	.246	.350	.401	.351	.407	.749	.695	.773	.804	.512	.673	.653	.687	.716	.658	.766	.593	.747	.577	.685	1.184	1.596		
F4	.243	.278	.217	.319	.229	.325	.373	.327	.378	.696	.646	.719	.748	.476	.626	.607	.639	.666	.612	.712	.551	.695	.537	.637	1.102	1.199	1.469	
F5	.068	.078	.061	.089	.064	.091	.105	.092	.106	.195	.181	.202	.210	.134	.176	.170	.179	.187	.172	.200	.155	.195	.151	.179	.309	.337	.313	1.296

Table 13. Both States 28 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Factor #1	C1	0.598	0.014	42.489	0.000
	C2	0.842	0.016	52.568	0.000
	C3	0.420	0.013	31.551	0.000
	C4	0.775	0.015	53.237	0.000
	C5	0.575	0.018	32.039	0.000
	C6	0.825	0.019	42.846	0.000
	C7	0.804	0.017	46.310	0.000
	C8	0.543	0.012	43.736	0.000
	C9	0.800	0.015	52.056	0.000
Factor #2	E1	0.617	0.013	47.169	0.000
	E2	0.685	0.013	51.070	0.000
	E3	0.808	0.015	54.276	0.000
	E4	0.781	0.020	39.599	0.000
	E6	0.641	0.013	47.722	0.000
	E7	0.665	0.012	57.062	0.000
	E8	0.675	0.013	52.511	0.000
	E11	0.658	0.016	40.755	0.000
	E12	0.701	0.013	54.114	0.000
	E13	0.616	0.013	46.640	0.000
Factor #3	P1	0.880	0.016	55.813	0.000
	P2	0.659	0.012	52.768	0.000
	P3	0.910	0.015	60.612	0.000
	P4	0.864	0.017	51.623	0.000
Factor #4	F1	0.489	0.014	35.536	0.000
	F2	0.989	0.014	70.327	0.000
	F3	1.027	0.014	74.412	0.000
	F4	0.888	0.014	61.553	0.000
	F5	0.522	0.018	28.649	0.000
Factor #2	Factor #1	0.733	0.009	78.116	0.000
Factor #3	Factor #1	0.742	0.010	75.804	0.000
Factor #3	Factor #2	0.864	0.007	130.647	0.000
Factor #4	Factor #1	0.531	0.013	40.818	0.000
Factor #4	Factor #2	0.767	0.008	93.282	0.000
Factor #4	Factor #3	0.614	0.012	51.794	0.000

Table 14. Both States 28 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.069	0.007	0.015	0.000
C2	0.088	0.009	0.018	0.001
C3	0.049	0.005	0.010	0.000
C4	0.100	0.010	0.021	0.001
C5	0.037	0.004	0.008	0.000
C6	0.051	0.005	0.011	0.000
C7	0.064	0.006	0.014	0.000
C8	0.082	0.008	0.017	0.001
C9	0.090	0.009	0.019	0.001
E1	0.008	0.055	0.030	0.008
E2	0.009	0.062	0.033	0.009
E3	0.009	0.063	0.033	0.009
E4	0.004	0.028	0.015	0.004
E6	0.008	0.055	0.029	0.008
E7	0.013	0.091	0.048	0.013
E8	0.010	0.069	0.037	0.010
E11	0.005	0.036	0.019	0.005
E12	0.010	0.072	0.038	0.011
E13	0.008	0.054	0.029	0.008
P1	0.015	0.025	0.149	0.001
P2	0.016	0.028	0.167	0.001
P3	0.019	0.032	0.190	0.001
P4	0.012	0.020	0.119	0.001
F1	0.000	0.007	0.001	0.033
F2	0.002	0.026	0.003	0.127
F3	0.002	0.034	0.004	0.168
F4	0.001	0.017	0.002	0.082
F5	0.000	0.004	0.000	0.019

Table 15. Both States 28 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E6	E7	E8	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	.976																											
C2	.504	1.391																										
C3	.251	.353	.789																									
C4	.464	.653	.325	1.151																								
C5	.344	.484	.241	.446	1.449																							
C6	.494	.695	.346	.640	.475	1.830																						
C7	.481	.677	.337	.623	.463	.664	1.537																					
C8	.325	.457	.228	.421	.312	.448	.437	.767																				
C9	.479	.674	.336	.620	.460	.660	.643	.435	1.275																			
E1	.271	.381	.190	.351	.260	.373	.364	.246	.362	.909																		
E2	.301	.423	.211	.389	.289	.415	.404	.273	.402	.423	.993																	
E3	.354	.498	.249	.459	.341	.489	.476	.322	.474	.499	.554	1.265																
E4	.343	.482	.240	.444	.329	.473	.460	.311	.458	.482	.536	.632	1.936															
E6	.281	.396	.197	.364	.270	.388	.378	.255	.376	.396	.440	.518	.501	.966														
E7	.292	.410	.205	.378	.280	.402	.392	.265	.390	.411	.456	.538	.520	.427	.791													
E8	.296	.416	.208	.383	.285	.408	.398	.269	.396	.417	.463	.546	.527	.433	.449	.921												
E11	.289	.406	.202	.374	.277	.398	.388	.262	.386	.406	.451	.532	.514	.422	.438	.444	1.299											
E12	.307	.432	.216	.398	.295	.424	.413	.279	.411	.433	.480	.567	.548	.450	.466	.473	.461	.952										
E13	.270	.380	.189	.350	.259	.372	.363	.245	.361	.380	.422	.497	.481	.395	.409	.415	.405	.431	.917									
P1	.391	.549	.274	.506	.376	.539	.525	.354	.522	.469	.521	.614	.594	.487	.505	.513	.500	.533	.468	1.386								
P2	.292	.411	.205	.379	.281	.403	.393	.265	.391	.351	.390	.460	.445	.365	.378	.384	.375	.399	.350	.579	.844							
P3	.404	.569	.283	.524	.389	.558	.543	.367	.541	.485	.539	.636	.615	.504	.523	.531	.518	.551	.484	.801	.599	1.325						
P4	.384	.540	.269	.497	.369	.529	.515	.348	.513	.461	.511	.603	.583	.479	.496	.504	.491	.523	.459	.760	.569	.786	1.499					
F1	.155	.219	.109	.201	.149	.214	.209	.141	.208	.232	.257	.303	.293	.241	.249	.253	.247	.263	.231	.264	.198	.273	.259	.884				
F2	.314	.442	.220	.407	.302	.433	.422	.285	.420	.468	.520	.613	.593	.487	.505	.512	.499	.532	.467	.534	.400	.553	.525	.483	1.317			
F3	.327	.459	.229	.423	.314	.450	.439	.296	.437	.487	.540	.637	.616	.506	.524	.532	.519	.553	.485	.555	.416	.574	.545	.502	1.016	1.321		
F4	.282	.397	.198	.366	.271	.390	.379	.256	.378	.421	.467	.551	.533	.437	.454	.460	.449	.478	.420	.480	.359	.497	.471	.434	.878	.913	1.262	
F5	.166	.234	.116	.215	.160	.229	.223	.151	.222	.247	.275	.324	.313	.257	.267	.271	.264	.281	.247	.282	.211	.292	.277	.255	.516	.537	.464	1.482

Table 16. Midwestern CW 31 Items- Standardized Coefficients

Factor	SIS Items	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Factor #1	C1	0.574	0.016	36.407	0.000
	C2	0.824	0.018	44.883	0.000
	C3	0.333	0.013	24.736	0.000
	C4	0.727	0.016	44.933	0.000
	C5	0.572	0.020	29.273	0.000
	C6	0.803	0.021	38.459	0.000
	C7	0.761	0.018	41.649	0.000
	C8	0.432	0.013	33.181	0.000
	C9	0.717	0.016	43.530	0.000
Factor #2	E1	0.385	0.011	34.893	0.000
	E2	0.442	0.013	34.229	0.000
	E3	0.625	0.013	47.969	0.000
	E4	0.785	0.022	36.153	0.000
	E5	0.651	0.013	49.402	0.000
	E6	0.691	0.014	50.275	0.000
	E7	0.483	0.011	45.344	0.000
	E8	0.472	0.013	37.375	0.000
	E9	0.609	0.012	49.649	0.000
	E10	0.493	0.012	42.518	0.000
	E11	0.618	0.017	35.996	0.000
	E12	0.477	0.012	38.236	0.000
	E13	0.414	0.012	34.634	0.000
Factor #3	P1	0.753	0.016	48.003	0.000
	P2	0.513	0.012	42.757	0.000
	P3	0.810	0.015	53.031	0.000
	P4	0.732	0.018	41.202	0.000
Factor #4	F1	0.363	0.014	26.492	0.000
	F2	0.825	0.013	65.974	0.000
	F3	0.870	0.013	66.419	0.000
	F4	0.669	0.014	46.158	0.000
	F5	0.540	0.020	27.027	0.000
Factor #2	Factor #1	0.719	0.011	64.395	0.000
Factor #3	Factor #1	0.778	0.011	72.923	0.000
Factor #3	Factor #2	0.770	0.010	76.878	0.000
Factor #4	Factor #1	0.503	0.016	31.986	0.000
Factor #4	Factor #2	0.664	0.012	56.592	0.000
Factor #4	Factor #3	0.462	0.017	27.822	0.000

Table 17. Midwestern CW 31 Items- Factor Score Coefficients

SIS Items	Factor #1	Factor #2	Factor #3	Factor #4
C1	0.068	0.003	0.017	0.001
C2	0.085	0.004	0.022	0.001
C3	0.048	0.002	0.012	0.001
C4	0.099	0.004	0.025	0.001
C5	0.041	0.002	0.010	0.000
C6	0.056	0.002	0.014	0.001
C7	0.074	0.003	0.019	0.001
C8	0.072	0.003	0.018	0.001
C9	0.089	0.004	0.023	0.001
E1	0.006	0.036	0.015	0.003
E2	0.005	0.030	0.012	0.002
E3	0.008	0.054	0.022	0.004
E4	0.003	0.019	0.008	0.002
E5	0.009	0.057	0.023	0.005
E6	0.009	0.057	0.023	0.005
E7	0.009	0.059	0.024	0.005
E8	0.005	0.035	0.014	0.003
E9	0.010	0.062	0.026	0.005
E10	0.007	0.047	0.020	0.004
E11	0.004	0.024	0.010	0.002
E12	0.006	0.037	0.015	0.003
E13	0.005	0.033	0.014	0.003
P1	0.023	0.010	0.172	-0.001
P2	0.024	0.011	0.175	-0.001
P3	0.031	0.014	0.232	-0.001
P4	0.015	0.007	0.111	-0.001
F1	0.001	0.002	-0.001	0.019
F2	0.008	0.015	-0.007	0.163
F3	0.008	0.015	-0.007	0.164
F4	0.002	0.004	-0.002	0.043
F5	0.001	0.001	-0.001	0.014

Table 18. Midwestern CW 31 Items- Estimated Covariance Correlations

	C1	C2	C3	C4	C5	C6	C7	C8	C9	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	P1	P2	P3	P4	F1	F2	F3	F4	F5
C1	.865																													
C2	.473	1.294																												
C3	.191	.275	.558																											
C4	.418	.599	.242	.999																										
C5	.329	.472	.191	.416	1.226																									
C6	.461	.662	.268	.584	.460	1.553																								
C7	.437	.627	.254	.553	.435	.611	1.234																							
C8	.248	.356	.144	.314	.247	.347	.328	.568																						
C9	.412	.591	.239	.521	.410	.576	.546	.310	1.026																					
E1	.159	.228	.092	.201	.159	.223	.211	.120	.199	.433																				
E2	.183	.262	.106	.231	.182	.256	.242	.137	.228	.170	.588																			
E3	.258	.371	.150	.327	.257	.361	.342	.194	.322	.241	.277	.699																		
E4	.324	.465	.188	.410	.323	.453	.429	.244	.405	.302	.347	.491	1.692																	
E5	.269	.386	.156	.340	.268	.376	.356	.202	.336	.251	.288	.407	.511	.727																
E6	.285	.409	.166	.361	.284	.399	.378	.214	.356	.266	.306	.432	.542	.450	.799															
E7	.200	.286	.116	.253	.199	.279	.264	.150	.249	.186	.214	.302	.379	.315	.334	.451														
E8	.195	.280	.113	.247	.194	.273	.258	.147	.244	.182	.209	.295	.370	.307	.326	.228	.580													
E9	.252	.361	.146	.319	.251	.352	.333	.189	.314	.235	.269	.381	.478	.397	.421	.294	.288	.632												
E10	.204	.292	.118	.258	.203	.285	.270	.153	.254	.190	.218	.308	.387	.321	.341	.238	.233	.300	.518											
E11	.255	.367	.148	.323	.254	.357	.338	.192	.319	.238	.274	.387	.485	.402	.427	.299	.292	.377	.305	1.054										
E12	.197	.283	.114	.249	.196	.276	.261	.148	.246	.184	.211	.298	.374	.311	.330	.231	.225	.291	.235	.295	.571									
E13	.171	.245	.099	.216	.170	.239	.226	.128	.213	.159	.183	.259	.324	.269	.286	.200	.195	.252	.204	.256	.197	.502								
P1	.337	.483	.195	.426	.335	.471	.446	.253	.420	.223	.256	.363	.455	.377	.400	.280	.274	.353	.286	.358	.277	.240	.970							
P2	.229	.329	.133	.290	.228	.321	.304	.172	.286	.152	.175	.247	.310	.257	.273	.191	.186	.241	.195	.244	.188	.163	.386	.533						
P3	.362	.520	.210	.458	.361	.507	.480	.272	.452	.240	.276	.390	.489	.406	.431	.301	.295	.380	.308	.386	.298	.258	.610	.416	.978					
P4	.327	.470	.190	.414	.326	.458	.434	.246	.409	.217	.249	.352	.442	.367	.389	.272	.266	.343	.278	.349	.269	.233	.552	.376	.593	1.143				
F1	.105	.151	.061	.133	.104	.147	.139	.079	.131	.093	.107	.151	.189	.157	.166	.116	.114	.147	.119	.149	.115	.100	.126	.086	.136	.123	.605			
F2	.239	.342	.138	.302	.238	.334	.316	.179	.298	.211	.242	.343	.430	.357	.378	.265	.259	.334	.270	.339	.261	.227	.287	.196	.309	.279	.299	.807		
F3	.252	.361	.146	.319	.251	.352	.333	.189	.314	.223	.256	.361	.454	.376	.399	.279	.273	.352	.285	.357	.276	.239	.303	.206	.326	.295	.316	.718	.890	
F4	.193	.278	.112	.245	.193	.271	.256	.145	.242	.171	.197	.278	.349	.289	.307	.215	.210	.271	.219	.275	.212	.184	.233	.159	.251	.227	.243	.552	.582	.838	
F5	.156	.224	.091	.197	.155	.218	.207	.117	.195	.138	.159	.224	.281	.233	.248	.173	.169	.218	.177	.222	.171	.148	.188	.128	.202	.183	.196	.445	.470	.361	1.288

Running Head: PSYCHOMETRIC EXAMINATION OF THE SIS

Appendix E: Item-Factor Loading Tables

Table 1. Midwestern CW 28 items- EFA specifying 4 factors Rotated Item Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.754	.402	.319	.388
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.726	.522	.495	.479
C9- I think that the current school initiatives are improving education for students in my school.	.706	.466	.407	.435
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.671	.474	.424	.390
F5- I think my school does a good job of including parents as team members in data-based decision making.	.665	.379	.510	.461
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.659	.537	.348	.387
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.635	.446	.374	.435
E11- I am involved in meetings where data results are discussed.	.627	.514	.520	.480
C1- I can summarize my schools' shared vision/mission.	.598	.411	.373	.378
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.593	.376		.327
C5- I have the technology and resources that I need to provide effective instruction.	.517			
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.505	.761	.539	.492
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.347	.706	.350	.391
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.578	.680	.375	.383
P3- I participate in professional development where I learn how to monitor students' progress.	.656	.680	.438	.465
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.569	.680	.589	.549
C3- I have a clear understanding of the State Standards for my grade/subject.	.373	.672	.322	.351
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.436	.653	.510	.601
P2- I participate in professional development where I learn to improve my instructional practices.	.556	.603	.326	.418
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.473	.592	.378	.473
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.448	.579	.482	.549
F3- I regularly communicate with families regarding student behavioral goals/progress.	.406	.457	.905	.469
F2- I regularly communicate with families regarding student academic goals/progress.	.414	.473	.888	.473
F4- I make informed decisions based on feedback from families.	.515	.468	.747	.553
E4- I review universal screening data at least three times a year for every student that I support.	.513	.491	.570	.455
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.481	.471	.466	.830

E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.490	.532	.473	.814
F1- I consider my students' backgrounds when planning instruction.	.479	.520	.499	.530

Table 2. Western EBISS 28 items- EFA specifying 4 factors Item Factor Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.740		.387	
C9- I think that the current tiered support model is improving education for students at my school.	.728	.372	.494	.308
C2- I receive school-wide academic and behavior data in usable and understandable formats.	.727	.333	.499	.324
C8- I feel that my administrators are committed to implementing tiered levels of academic supports.	.706		.394	
E11- I am involved in meetings where data results are discussed and problem solving occurs.	.655	.461	.585	.455
C7- I am involved in action planning to implement tiered supports with other staff and administrators at my school.	.601	.344	.503	.427
F5- I think that my school does a good job of including parents as team members in data-based decision making.	.577		.392	
C1- I can summarize my school's shared vision/mission.	.522		.311	
C6- I have the time necessary to analyze student data and problem solve with my colleagues.	.473		.467	
C5- I have the technology and resources that I need to provide effective instruction.	.407			
E1- My instruction intentionally addresses the Common Core State Standards for my grade/subject.		.844	.461	.493
E2- I am able to differentiate instruction according to student needs while addressing the Common Core State Standards.	.353	.772	.590	.498
P1- I participate in professional development where I learn how to develop curricular plans that address the Common Core State Standards.	.379	.765	.535	.442
C3- I have a clear understanding of the Common Core State Standards.	.350	.652	.422	.428
P2- I participate in professional development where I learn strategies to improve my instructional practices.	.437	.640	.597	.421
E6- I evaluate the effectiveness of my instruction based on assessment data.	.362	.609	.581	.418
E12- When I'm concerned about a student's academic success, I collaborate with a team to identify interventions.	.528	.470	.761	.465
E7- I adapt the environment, curriculum, and instruction based on each student's academic data.	.379	.651	.754	.491
E8- I adapt the environment, curriculum, and instruction based on each student's behavior data.	.457	.548	.750	.487
E13- When I'm concerned about a student's behavioral success, I collaborate with a team to identify interventions.	.552	.357	.702	.409
E3- I frequently monitor the progress of students receiving Tier 2 and Tier 3 interventions.	.534	.519	.691	.594
P3- I participate in professional development where I learn how to monitor students' progress and use progress monitoring data.	.548	.590	.680	.478
E4- I review universal screening data at least three times a year for every student I support.	.386	.585	.610	.511

F1- I consider my students' background when planning instruction.		.552	.602	.531
P4- I receive coaching/mentoring to help me implement tiered levels of academic support.	.549	.494	.558	.319
F3- I regularly communicate with families regarding student behavior goals/progress.	.322	.528	.550	.928
F4- I make informed decisions based on feedback from families.	.314	.534	.583	.856
F2- I regularly communicate with families regarding student academic goals/progress.		.572	.487	.762

Table 3. SW-PBS/SWPBIS 28 items- EFA specifying 4 factors Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.775	.414	.474	.500
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.760	.369	.599	.560
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.750		.445	.468
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.701	.443	.597	.549
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.670	.414	.523	.385
P2- I participate in professional development where I learn to improve my instructional practices.	.662	.378	.473	.346
F1- I consider my students' backgrounds when planning instruction.	.606		.450	.519
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.587	.580	.567	.326
C3- I have a clear understanding of the State Standards for my grade/subject.	.563		.427	.356
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.340	.790	.446	
C9- I think that the current school initiatives are improving education for students in my school.	.474	.718	.630	.356
F5- I think my school does a good job of including parents as team members in data-based decision making.	.358	.640	.487	.385
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.309	.616	.474	
C1- I can summarize my schools' shared vision/mission.	.333	.559	.451	
C5- I have the technology and resources that I need to provide effective instruction.		.485		
E11- I am involved in meetings where data results are discussed.	.466	.545	.747	.411
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.450	.663	.730	.366
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.599	.481	.714	.583
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.623	.412	.683	.537
P3- I participate in professional development where I learn how to monitor students' progress.	.613	.532	.677	.403
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.532	.471	.677	.538
E4- I review universal screening data at least three times a year for every student that I support.	.497	.380	.655	.448

C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.388	.570	.622	.365
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.600	.401	.618	.449
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.414	.550	.556	.358
F3- I regularly communicate with families regarding student behavioral goals/progress.	.535	.331	.562	.898
F2- I regularly communicate with families regarding student academic goals/progress.	.614	.317	.532	.803
F4- I make informed decisions based on feedback from families.	.530	.357	.535	.797

Table 4. RTI 28 items- EFA specifying 4 factors Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
F3- I regularly communicate with families regarding student behavioral goals/progress.	.820	.548		.393
F4- I make informed decisions based on feedback from families.	.794	.553		.362
F2- I regularly communicate with families regarding student academic goals/progress.	.733	.530		
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.731	.478	.377	.608
E4- I review universal screening data at least three times a year for every student that I support.	.685	.630	.386	.373
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.663	.489	.465	.638
F1- I consider my students' backgrounds when planning instruction.	.638	.605	.322	.533
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.519	.513	.350	.444
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.602	.799		.372
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.601	.764	.365	.589
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.579	.752		.474
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.550	.663		.493
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.624	.661	.438	.654
C3- I have a clear understanding of the State Standards for my grade/subject.	.440	.635	.504	
P2- I participate in professional development where I learn to improve my instructional practices.	.446	.633	.461	.607
E11- I am involved in meetings where data results are discussed.	.449	.444	.775	.457
C9- I think that the current school initiatives are improving education for students in my school.	.325		.772	.570
C2- I receive school-wide academic and behavioral data in usable and understandable formats.		.327	.767	.385
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).			.750	.491
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.416	.344	.631	.606
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.463	.445	.607	.561

E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.594	.429	.470	.711
P3- I participate in professional development where I learn how to monitor students' progress.	.631	.643	.430	.651
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.378	.376		.642
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.418	.474	.390	.624
F5- I think my school does a good job of including parents as team members in data-based decision making.			.407	.594
C1- I can summarize my schools' shared vision/mission.		.383	.406	.463
C5- I have the technology and resources that I need to provide effective instruction.				.445

Table 5. Both States 28 items- EFA specifying 4 factors Rotated Item Factor Structure Matrix

	Factor			
School Implementation Scale Items	1	2	3	4
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.768	.495	.589	
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.747	.475	.488	
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.739	.330	.446	.385
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.708	.533	.576	
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.703	.481	.413	.525
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.687	.546	.598	.476
P3- I participate in professional development where I learn how to monitor students' progress.	.672	.595	.465	.579
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.664	.574	.578	
P2- I participate in professional development where I learn to improve my instructional practices.	.663	.501	.404	.399
F1- I consider my students' backgrounds when planning instruction.	.611	.435	.543	
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.611	.484	.510	.525
C3- I have a clear understanding of the State Standards for my grade/subject.	.601	.370	.363	.419
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.598	.596	.560	
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.396	.771		.349
C9- I think that the current school initiatives are improving education for students in my school.	.493	.715	.392	.460
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.431	.680	.388	.632
F5- I think my school does a good job of including parents as team members in data-based decision making.	.410	.671	.447	
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.405	.633		.355

C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.446	.621	.397	.514
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.590	.620	.381	.440
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.452	.607	.356	.392
C1- I can summarize my schools' shared vision/mission.	.392	.580	.330	.344
C5- I have the technology and resources that I need to provide effective instruction.		.524		
F3- I regularly communicate with families regarding student behavioral goals/progress.	.558	.428	.899	.389
F2- I regularly communicate with families regarding student academic goals/progress.	.579	.404	.823	.367
F4- I make informed decisions based on feedback from families.	.565	.476	.794	.336
E11- I am involved in meetings where data results are discussed.	.475	.598	.451	.662
E4- I review universal screening data at least three times a year for every student that I support.	.477	.416	.484	.531

Table 6. Midwestern CW 28 items- EFA Rotated Item Structure Matrix

	Factor					
School Implementation Scale Items	1	2	3	4	5	6
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.798	.334	.330	.507	.401	
C9- I think that the current school initiatives are improving education for students in my school.	.698	.411	.397	.522	.452	.327
F5- I think my school does a good job of including parents as team members in data-based decision making.	.683	.520	.304	.474	.467	
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.655	.468	.420	.575	.495	.567
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.635	.422	.391	.534	.410	.368
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.606	.363	.386	.472	.447	.396
C1- I can summarize my schools' shared vision/mission.	.605	.383	.362	.452	.392	
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.583			.513	.351	
C5- I have the technology and resources that I need to provide effective instruction.	.571			.316		
F3- I regularly communicate with families regarding student behavioral goals/progress.	.375	.889	.424	.409	.478	
F2- I regularly communicate with families regarding student academic goals/progress.	.376	.876	.430	.435	.485	
F4- I make informed decisions based on feedback from families.	.504	.757	.412	.478	.561	
E4- I review universal screening data at least three times a year for every student that I support.	.443	.552	.430	.469	.471	.431
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.319	.380	.800	.482	.433	
C3- I have a clear understanding of the State Standards for my grade/subject.	.358	.352	.757	.474	.390	
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.426	.559	.712	.650	.539	.313
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.491	.585	.638	.593	.579	.439

E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.493	.440	.571	.553	.518	
P3- I participate in professional development where I learn how to monitor students' progress.	.548	.462	.524	.793	.513	.365
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.474	.403	.553	.744	.433	.310
P2- I participate in professional development where I learn to improve my instructional practices.	.490	.373	.466	.724	.470	
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.603	.376	.391	.676	.426	
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.446	.485	.477	.510	.823	
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.459	.482	.421	.469	.802	
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.435	.577	.621	.592	.648	
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.461	.558	.530	.580	.600	
F1- I consider my students' backgrounds when planning instruction.	.493	.537	.495	.484	.553	
E11- I am involved in meetings where data results are discussed.	.533	.490	.419	.543	.494	.582

Table 7. Western EBISS 28 items- EFA Rotated Item Structure Matrix

School Implementation Scale Items	Factor				
	1	2	3	4	5
C2- I receive school-wide academic and behavior data in usable and understandable formats.	.738	.349	.344	.435	.467
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.737			.310	.399
C9- I think that the current tiered support model is improving education for students at my school.	.732	.391	.327	.424	.480
C8- I feel that my administrators are committed to implementing tiered levels of academic supports.	.702			.315	.407
E11- I am involved in meetings where data results are discussed and problem solving occurs.	.669	.478	.477	.517	.540
C7- I am involved in action planning to implement tiered supports with other staff and administrators at my school.	.606	.368	.442	.409	.497
F5- I think that my school does a good job of including parents as team members in data-based decision making.	.574			.303	.411
P4- I receive coaching/mentoring to help me implement tiered levels of academic support.	.558	.521	.343	.472	.542
C1- I can summarize my school's shared vision/mission.	.519				.316
C5- I have the technology and resources that I need to provide effective instruction.	.403				
E1- My instruction intentionally addresses the Common Core State Standards for my grade/subject.		.833	.511	.508	.371
P1- I participate in professional development where I learn how to develop curricular plans that address the Common Core State Standards.	.393	.798	.461	.492	.499
E2- I am able to differentiate instruction according to student needs while addressing the Common Core State Standards.	.378	.768	.522	.608	.493
P2- I participate in professional development where I learn strategies to improve my instructional practices.	.456	.654	.446	.558	.536
C3- I have a clear understanding of the Common Core State Standards.	.367	.646	.444	.446	.349

F3- I regularly communicate with families regarding student behavior goals/progress.	.342	.534	.922	.512	.481
F4- I make informed decisions based on feedback from families.	.336	.541	.857	.544	.512
F2- I regularly communicate with families regarding student academic goals/progress.		.564	.763	.508	.386
E7- I adapt the environment, curriculum, and instruction based on each student's academic data.	.424	.630	.524	.924	.555
E8- I adapt the environment, curriculum, and instruction based on each student's behavior data.	.499	.532	.522	.811	.596
E4- I review universal screening data at least three times a year for every student I support.	.420	.579	.538	.630	.481
E6- I evaluate the effectiveness of my instruction based on assessment data.	.395	.600	.446	.622	.449
F1- I consider my students' background when planning instruction.	.301	.555	.550	.581	.518
E12- When I'm concerned about a student's academic success, I collaborate with a team to identify interventions.	.529	.521	.490	.575	.848
E13- When I'm concerned about a student's behavioral success, I collaborate with a team to identify interventions.	.551	.405	.432	.502	.811
E3- I frequently monitor the progress of students receiving Tier 2 and Tier 3 interventions.	.551	.541	.614	.600	.648
P3- I participate in professional development where I learn how to monitor students' progress and use progress monitoring data.	.564	.619	.505	.586	.642
C6- I have the time necessary to analyze student data and problem solve with my colleagues.	.473	.314		.348	.492

Table 8. SW-PBS/SWPBIS 28 items- EFA Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.738	.409	.445	.496
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.673	.335	.519	.512
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.712		.420	.466
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.671	.439	.562	.543
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.787	.471	.599	.453
P2- I participate in professional development where I learn to improve my instructional practices.	.621	.343	.421	.319
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.737	.710	.687	.415
F1- I consider my students' backgrounds when planning instruction.	.561		.410	.494
C3- I have a clear understanding of the State Standards for my grade/subject.	.488		.376	.323
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.367	.838	.482	
C9- I think that the current school initiatives are improving education for students in my school.	.527	.780	.702	.403
F5- I think my school does a good job of including parents as team members in data-based decision making.	.433	.765	.573	.461
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.544	.711	.705	.467
C5- I have the technology and resources that I need to provide effective instruction.	.324	.557	.327	

C8- I feel that my administrators are committed to implementing evidence-based instructional practices.		.548	.422	
C1- I can summarize my schools' shared vision/mission.	.319	.533	.438	
E4- I review universal screening data at least three times a year for every student that I support.	.652	.498	.848	.597
E11- I am involved in meetings where data results are discussed.	.525	.601	.847	.472
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.512	.738	.833	.425
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.705	.566	.823	.691
P3- I participate in professional development where I learn how to monitor students' progress.	.737	.615	.788	.480
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.627	.424	.656	.546
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.530	.478	.647	.543
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.484	.679	.760	.453
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.557	.368	.577	.429
F3- I regularly communicate with families regarding student behavioral goals/progress.	.595	.367	.613	.989
F2- I regularly communicate with families regarding student academic goals/progress.	.672	.351	.567	.879
F4- I make informed decisions based on feedback from families.	.566	.386	.559	.851

Table 9. RTI 28 items- EFA Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor				
	1	2	3	4	5
F3- I regularly communicate with families regarding student behavioral goals/progress.	.822		.390	.499	.460
F4- I make informed decisions based on feedback from families.	.793		.351	.487	.497
F2- I regularly communicate with families regarding student academic goals/progress.	.735			.486	.433
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.733	.372	.617	.428	.424
E4- I review universal screening data at least three times a year for every student that I support.	.691	.379	.361	.590	.493
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.662	.462	.642	.439	.440
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.514	.354	.424	.445	.491
C9- I think that the current school initiatives are improving education for students in my school.		.798	.555		.372
E11- I am involved in meetings where data results are discussed.	.453	.765	.458	.445	.316
C2- I receive school-wide academic and behavioral data in usable and understandable formats.		.758	.385	.336	
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).		.754	.484		
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.406	.638	.597		.387
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.463	.603	.560	.418	.370

E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.591	.466	.724	.388	.393
P3- I participate in professional development where I learn how to monitor students' progress.	.635	.420	.655	.619	.494
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.368	.324	.635	.317	.398
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.412	.390	.617	.432	.427
F5- I think my school does a good job of including parents as team members in data-based decision making.		.412	.590		
C1- I can summarize my schools' shared vision/mission.		.397	.475	.399	
C5- I have the technology and resources that I need to provide effective instruction.			.446		
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.616	.335	.612	.830	.446
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.614		.356	.774	.556
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.581		.446	.685	.621
C3- I have a clear understanding of the State Standards for my grade/subject.	.451	.493		.623	.434
P2- I participate in professional development where I learn to improve my instructional practices.	.446	.453	.599	.614	.485
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.603	.458	.612	.511	.824
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.529	.301	.435	.535	.821
F1- I consider my students' backgrounds when planning instruction.	.628	.336	.493	.471	.710

Table 10. Both States 28 items- EFA Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor			
	1	2	3	4
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.768	.495	.589	
E2- I am able to differentiate instruction according ro student needs while addressing the State Standards.	.747	.475	.488	
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.739	.330	.446	.385
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.708	.533	.576	
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.703	.481	.413	.525
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.687	.546	.598	.476
P3- I participate in professional development where I learn how to monitor students' progress.	.672	.595	.465	.579
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.664	.574	.578	
P2- I participate in professional development where I learn to improve my instructional practices.	.663	.501	.404	.399
F1- I consider my students' backgrounds when planning instruction.	.611	.435	.543	
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.611	.484	.510	.525

C3- I have a clear understanding of the State Standards for my grade/subject.	.601	.370	.363	.419
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.598	.596	.560	
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.396	.771		.349
C9- I think that the current school initiatives are improving education for students in my school.	.493	.715	.392	.460
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.431	.680	.388	.632
F5- I think my school does a good job of including parents as team members in data-based decision making.	.410	.671	.447	
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.405	.633		.355
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.446	.621	.397	.514
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.590	.620	.381	.440
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.452	.607	.356	.392
C1- I can summarize my schools' shared vision/mission.	.392	.580	.330	.344
C5- I have the technology and resources that I need to provide effective instruction.		.524		
F3- I regularly communicate with families regarding student behavioral goals/progress.	.558	.428	.899	.389
F2- I regularly communicate with families regarding student academic goals/progress.	.579	.404	.823	.367
F4- I make informed decisions based on feedback from families.	.565	.476	.794	.336
E11- I am involved in meetings where data results are discussed.	.475	.598	.451	.662
E4- I review universal screening data at least three times a year for every student that I support.	.477	.416	.484	.531

Table 11. Midwestern CW 37 items- EFA Rotated Factor Structure Matrix

	Factor						
School Implementation Scale Items	1	2	3	4	5	6	7
E9- I modify my instructional practices based on students' common formative assessment data.	.782	.394		.492	.449	.491	.393
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.771	.408		.500	.416	.560	
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.738	.395		.484	.582	.402	.530
E5- I review formative assessment data for every student that I support.	.716	.357		.504	.524	.452	.584
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.712	.457	.307	.524	.498	.475	.643
E10- Based on assessment results, I re-teach information that students have not mastered.	.707	.340		.427	.403	.466	.322
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.656				.434		.364
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.644	.443		.490	.425	.518	
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.615	.477		.364	.439	.416	
C3- I have a clear understanding of the State Standards for my grade/subject.	.602	.342			.443		.326

F1- I consider my students' backgrounds when planning instruction.	.562	.479		.490	.376	.481	
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.351	.793	.453	.318	.508	.364	
C9- I think that the current school initiatives are improving education for students in my school.	.437	.696	.389	.379	.477	.395	.392
F5- I think my school does a good job of including parents as team members in data-based decision making.	.378	.686	.409	.523	.428	.447	
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.462	.649	.422	.439	.538	.441	.621
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.406	.633	.334	.401	.526	.358	.420
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.410	.604	.344	.331	.435	.399	.449
C1- I can summarize my schools' shared vision/mission.	.382	.597	.323	.363	.431	.348	
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.315	.579	.334		.504	.328	
C5- I have the technology and resources that I need to provide effective instruction.		.566					
Mo3- The CDTs in my building have been trained to collect and analyze data to inform instruction.		.369	.753				
Mo1- My building has a collaborative data team (CDT) that meet regularly (at least once a month).			.697				
Mo7- The CDTs in my building develop and administer Common Formative Assessments and use the results to inform instruction.		.356	.696				
Mo6- All teachers have been trained to implement the identified effective teaching practices.		.459	.574		.403		
Mo5- My school has identified at least three effective teaching practices to implement in classroom instruction.		.403	.561		.357		
Mo2- The CDT structure in my building include representatives from all teaching roles (i.e., regular, education, special education, special classes [music, art, PE], etc.).		.337	.558				
F3- I regularly communicate with families regarding student behavioral goals/progress.	.503	.361		.906	.319	.426	.381
F2- I regularly communicate with families regarding student academic goals/progress.	.516	.361		.891	.344	.431	.389
F4- I make informed decisions based on feedback from families.	.507	.493		.756	.383	.522	.319
P3- I participate in professional development where I learn how to monitor students' progress.	.574	.538	.391	.418	.789	.437	.484
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.549	.469		.353	.770	.336	.426
P2- I participate in professional development where I learn to improve my instructional practices.	.513	.483		.326	.705	.410	
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.444	.602	.386	.348	.671	.376	.328
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.553	.430		.444	.426	.821	.391

E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.501	.450		.451	.384	.819	.305
E11- I am involved in meetings where data results are discussed.	.460	.517	.386	.461	.502	.455	.665
E4- I review universal screening data at least three times a year for every student that I support.	.504	.429	.310	.526	.398	.403	.539

Table 12. Western EBISS 42 items- EFA Rotated Factor Structure Matrix

	Factor						
School Implementation Scale Items	1	2	3	4	5	6	7
T7- I think my school does a good job of addressing the behavior needs of students in tier 2 (small group).	.844		.537		.527		.362
T4- I think my school does a good job of addressing the academic needs of students in tier 2 (small group).	.789		.447		.447	.314	.402
T8- I think my school does a good job of addressing the behavior needs of students in tier 3 (intensive).	.778		.541		.444		.391
T5- I think my school does a good job of addressing the academic needs of students in tier 3 (intensive).	.722		.480		.414		.425
T6- I think my school does a good job of addressing the behavior needs of students in tier 1 (universal).	.703		.399		.509		
T3- I think my school does a good job of addressing the academic needs of students in tier 1 (universal).	.700	.329	.393		.469	.344	.302
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.672		.484		.545		.510
F5- I think that my school does a good job of including parents as team members in data-based decision making.	.594		.469		.413		.449
C5- I have the technology and resources that I need to provide effective instruction.	.440						.322
E1- My instruction intentionally addresses the Common Core State Standards for my grade/subject.		.859		.527	.333	.513	.322
E2- I am able to differentiate instruction according to student needs while addressing the Common Core State Standards.		.779	.433	.526	.397	.599	.385
P1- I participate in professional development where I learn how to develop curricular plans that address the Common Core State Standards.		.772	.411	.470	.388	.515	.494
C3- I have a clear understanding of the Common Core State Standards.		.659		.424	.429	.451	
P2- I participate in professional development where I learn strategies to improve my instructional practices.	.310	.608	.452	.427	.446	.573	.523
E13- When I'm concerned about a student's behavioral success, I collaborate with a team to identify interventions.	.473	.357	.837	.391	.505	.476	.474

T12- I feel that the team that addresses behavioral needs provides valuable feedback and makes informed decisions.	.656		.808		.525	.307	.472
T11- I feel that the team that addresses academic needs provides valuable feedback and makes informed decisions.	.543	.363	.789	.324	.559	.495	.559
E12- When I'm concerned about a student's academic success, I collaborate with a team to identify interventions.	.373	.469	.781	.446	.527	.589	.517
T14- I regularly see students more between tiers of support as their behavioral needs change.	.666		.775	.319	.546	.405	.481
T13- I regularly see students more between tiers of support as their academic needs change.	.522	.400	.712	.342	.521	.509	.544
E3- I frequently monitor the progress of students receiving Tier 2 and Tier 3 interventions.	.413	.510	.627	.580	.598	.597	.435
C6- I have the time necessary to analyze student data and problem solve with my colleagues.	.400		.500		.386	.339	.442
F3- I regularly communicate with families regarding student behavior goals/progress.		.518	.410	.916	.427	.523	.328
F4- I make informed decisions based on feedback from families.		.530	.436	.835	.425	.556	
F2- I regularly communicate with families regarding student academic goals/progress.		.568		.828	.345	.528	.313
Or1- I regularly communicate with families regarding student progress in meeting the Common Core State Standards.		.624	.351	.740	.359	.489	.383
C7- I am involved in action planning to implement tiered supports with other staff and administrators at my school.	.388	.323	.489	.368	.737	.399	.401
C8- I feel that my administrators are committed to implementing tiered levels of academic supports.	.649		.465		.726		.422
E11- I am involved in meetings where data results are discussed and problem solving occurs.	.438	.415	.510	.406	.700	.534	.524
T1- I have a clear understanding of the phrase "tiered levels of academic and behavior supports."	.412	.386	.398	.335	.694	.414	.318
T2- I feel that my administrators are committed to implementing tiered levels of behavior supports.	.690		.525		.692		.395
C9- I think that the current tiered support model is improving education for students at my school.	.660	.346	.532		.679	.398	.477
C2- I receive school-wide academic and behavior data in usable and understandable formats.	.568		.508		.656	.429	.541
C1- I can summarize my school's shared vision/mission.	.424		.360		.493		.330
E7- I adapt the environment, curriculum, and instruction based on each student's academic data.		.598	.487	.508	.433	.866	.425
E8- I adapt the environment, curriculum, and instruction based on each student's behavior data.	.380	.491	.561	.494	.476	.771	.457
E4- I review universal screening data at least three times a year for every student I support.		.546	.394	.501	.518	.665	.419
E6- I evaluate the effectiveness of my instruction based on assessment data.		.578	.380	.420	.433	.638	.359
P3- I participate in professional development where I learn how to monitor students' progress and use progress monitoring data.	.350	.561	.557	.475	.572	.619	.604

F1- I consider my students' background when planning instruction.		.548	.441	.525	.355	.591	
P4- I receive coaching/mentoring to help me implement tiered levels of academic support.	.409	.458	.498	.335	.483	.474	.850
T10- I receive coaching/mentoring to help me implement tiered levels of behavior support.	.515	.338	.595	.321	.497	.390	.803

Table 13. Midwestern CW 31 items- EFA specifying 4 factors Rotated Item Factor Structure Matrix

School Implementation Scale Items	Factor 1	Factor 2	Factor 3	Factor 4
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.777	.334	.393	
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.699	.337	.621	.504
C9- I think that the current school initiatives are improving education for students in my school.	.699	.380	.482	.405
F5- I think my school does a good job of including parents as team members in data-based decision making.	.679	.382	.368	.520
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.658	.326	.509	.418
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.640	.389	.545	.327
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.622	.336	.489	.379
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	.598		.370	
C1- I can summarize my schools' shared vision/mission.	.598	.346	.407	.368
C5- I have the technology and resources that I need to provide effective instruction.	.534			
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.395	.801	.540	.479
E9- I modify my instructional practices based on students' common formative assessment data.	.406	.717	.653	.503
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.432	.693	.448	.452
E10- Based on assessment results, I re-teach information that students have not mastered.	.349	.665	.577	.433
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.459	.625	.469	.334
F1- I consider my students' backgrounds when planning instruction.	.468	.572	.438	.485
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.476	.556	.518	.493
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.477	.534	.447	.486
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.435	.601	.763	.498
E5- I review formative assessment data for every student that I support.	.411	.576	.751	.544
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.502	.566	.746	.582
P3- I participate in professional development where I learn how to monitor students' progress.	.612	.470	.718	.411
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.534	.439	.682	.333
E11- I am involved in meetings where data results are discussed.	.592	.327	.628	.536
E1- My instruction intentionally addresses the State Standards for my grade/subject.		.557	.606	
P2- I participate in professional development where I learn to improve my instructional practices.	.530	.473	.568	

C3- I have a clear understanding of the State Standards for my grade/subject.	.339	.511	.565	
F3- I regularly communicate with families regarding student behavioral goals/progress.	.374	.451	.456	.877
F2- I regularly communicate with families regarding student academic goals/progress.	.378	.461	.476	.863
F4- I make informed decisions based on feedback from families.	.501	.502	.440	.752
E4- I review universal screening data at least three times a year for every student that I support.	.467	.389	.570	.578

Table 14. Midwestern CW 31 items- EFA Rotated Item Factor Structure Matrix

	Factor					
School Implementation Scale Items	1	2	3	4	5	6
E9- I modify my instructional practices based on students' common formative assessment data.	.771	.391	.493	.473	.505	.480
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	.770	.404	.506	.444	.565	
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	.728	.386	.478	.590	.423	.603
E10- Based on assessment results, I re-teach information that students have not mastered.	.700	.336	.429	.425	.477	.401
E5- I review formative assessment data for every student that I support.	.697	.352	.499	.541	.470	.651
E1- My instruction intentionally addresses the State Standards for my grade/subject.	.663			.442		.411
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	.647	.442	.496	.447	.519	
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	.622	.473	.367	.454	.423	
C3- I have a clear understanding of the State Standards for my grade/subject.	.610	.327		.442		.376
F1- I consider my students' backgrounds when planning instruction.	.560	.476	.489	.382	.488	
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	.345	.792		.493	.375	.329
C9- I think that the current school initiatives are improving education for students in my school.	.423	.698	.367	.471	.408	.452
F5- I think my school does a good job of including parents as team members in data-based decision making.	.368	.685	.510	.419	.454	.339
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	.392	.632	.388	.518	.371	.472
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	.393	.605		.431	.411	.497
C1- I can summarize my schools' shared vision/mission.	.375	.596	.353	.425	.358	.350
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.		.582		.506	.334	
C5- I have the technology and resources that I need to provide effective instruction.		.566				
F3- I regularly communicate with families regarding student behavioral goals/progress.	.488	.359	.908	.327	.434	.444

F2- I regularly communicate with families regarding student academic goals/progress.	.500	.359	.893	.354	.439	.453
F4- I make informed decisions based on feedback from families.	.494	.493	.754	.389	.528	.393
P3- I participate in professional development where I learn how to monitor students' progress.	.559	.531	.405	.794	.452	.557
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	.540	.457	.342	.766	.355	.492
P2- I participate in professional development where I learn to improve my instructional practices.	.510	.478	.321	.727	.418	.333
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	.434	.597	.335	.671	.386	.398
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	.540	.430	.441	.443	.831	.448
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	.491	.451	.449	.396	.828	.365
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	.691	.454	.517	.510	.494	.706
E11- I am involved in meetings where data results are discussed.	.430	.527	.454	.516	.465	.679
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	.434	.653	.423	.537	.456	.669
E4- I review universal screening data at least three times a year for every student that I support.	.480	.427	.516	.397	.420	.602

Appendix F: SIS Alpha if Item Deleted Tables

Table 1. Midwestern State CW 28 items- Item-Total Statistics (Alpha for Scale=0.939)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C1- I can summarize my schools' shared vision/mission.	112.19	244.947	0.547	0.382	0.937
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	112.46	236.450	0.686	0.569	0.935
C3- I have a clear understanding of the State Standards for my grade/subject.	111.97	249.665	0.488	0.514	0.938
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	112.35	241.880	0.607	0.562	0.936
C5- I have the technology and resources that I need to provide effective instruction.	112.56	245.882	0.422	0.337	0.939
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	113.08	237.649	0.587	0.427	0.937
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	112.48	239.346	0.616	0.438	0.936
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	111.86	249.368	0.496	0.378	0.938
C9- I think that the current school initiatives are improving education for students in my school.	112.31	240.753	0.635	0.476	0.936
E1- My instruction intentionally addresses the State Standards for my grade/subject.	111.78	251.032	0.493	0.550	0.938
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	112.06	247.933	0.548	0.409	0.937
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	111.98	244.150	0.647	0.561	0.936
E4- I review universal screening data at least three times a year for every student that I support.	112.51	236.820	0.580	0.437	0.937
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	112.10	242.033	0.680	0.549	0.936
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	111.79	248.899	0.585	0.531	0.937
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	111.90	247.759	0.559	0.477	0.937
E11- I am involved in meetings where data results are discussed.	112.08	240.339	0.639	0.518	0.936

E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	111.79	247.219	0.588	0.601	0.937
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	111.78	248.673	0.563	0.582	0.937
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	112.15	242.085	0.610	0.539	0.936
P2- I participate in professional development where I learn to improve my instructional practices.	111.79	248.059	0.572	0.465	0.937
P3- I participate in professional development where I learn how to monitor students' progress.	112.19	239.983	0.678	0.591	0.936
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	112.39	240.506	0.606	0.459	0.937
F1- I consider my students' backgrounds when planning instruction.	112.00	247.300	0.566	0.387	0.937
F2- I regularly communicate with families regarding student academic goals/progress.	112.05	244.859	0.572	0.776	0.937
F3- I regularly communicate with families regarding student behavioral goals/progress.	112.08	244.271	0.562	0.782	0.937
F4- I make informed decisions based on feedback from families.	112.17	243.326	0.616	0.572	0.936
F5- I think my school does a good job of including parents as team members in data-based decision making.	112.63	239.095	0.609	0.487	0.937

Table 2. Western State EBISS 28 Items- Item-Total Statistics (Alpha for Scale =0.937)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C1- I can summarize my school's shared vision/mission.	98.41	369.363	.403	.261	.937
C2- I receive school-wide academic and behavior data in usable and understandable formats.	98.61	359.253	.588	.553	.935
C3- I have a clear understanding of the Common Core State Standards.	98.22	364.845	.546	.451	.935
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	98.67	367.452	.461	.528	.936
C5- I have the technology and resources that I need to provide effective instruction.	98.80	371.850	.279	.252	.938
C6- I have the time necessary to analyze student data and problem solve with my colleagues.	99.61	363.171	.449	.311	.936
C7- I am involved in action planning to implement tiered supports with other staff and administrators at my school.	98.79	356.452	.565	.447	.935

C8- I feel that my administrators are committed to implementing tiered levels of academic supports.	98.13	365.705	.510	.453	.936
C9- I think that the current tiered support model is improving education for students at my school.	98.74	361.510	.588	.512	.935
E1- My instruction intentionally addresses the Common Core State Standards for my grade/subject.	98.17	359.803	.571	.674	.935
E2- I am able to differentiate instruction according to student needs while addressing the Common Core State Standards.	98.47	358.193	.642	.638	.934
E3- I frequently monitor the progress of students receiving Tier 2 and Tier 3 interventions.	98.71	352.266	.693	.556	.933
E4- I review universal screening data at least three times a year for every student I support.	98.36	352.300	.615	.527	.934
E6- I evaluate the effectiveness of my instruction based on assessment data.	98.20	363.964	.575	.457	.935
E7- I adapt the environment, curriculum, and instruction based on each student's academic data.	98.17	361.018	.658	.693	.934
E8- I adapt the environment, curriculum, and instruction based on each student's behavior data.	98.29	360.439	.654	.644	.934
E11- I am involved in meetings where data results are discussed and problem solving occurs.	98.32	354.847	.656	.564	.934
E12- When I'm concerned about a student's academic success, I collaborate with a team to identify interventions.	98.25	356.363	.665	.626	.934
E13- When I'm concerned about a student's behavioral success, I collaborate with a team to identify interventions.	98.38	358.845	.607	.610	.934
P1- I participate in professional development where I learn how to develop curricular plans that address the Common Core State Standards.	98.53	353.909	.627	.599	.934
P2- I participate in professional development where I learn strategies to improve my instructional practices.	98.01	359.807	.637	.494	.934
P3- I participate in professional development where I learn how to monitor students' progress and use progress monitoring data.	98.65	351.653	.703	.573	.933
P4- I receive coaching/mentoring to help me implement tiered levels of academic support.	98.80	356.443	.586	.434	.935
F1- I consider my students' background when planning instruction.	98.04	364.700	.550	.445	.935
F2- I regularly communicate with families regarding student academic goals/progress.	98.70	358.754	.552	.637	.935
F3- I regularly communicate with families regarding student behavior goals/progress.	98.65	355.452	.609	.759	.934
F4- I make informed decisions based on feedback from families.	98.53	356.207	.614	.708	.934

F5- I think that my school does a good job of including parents as team members in data-based decision making.	99.09	367.087	.421	.322	.937

Table 3. SW-PBS/SWPBIS 28 Items- Item-Total Statistics (Alpha for Scale=0.942)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	107.44	332.735	.483	.383	.941
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	107.73	326.179	.632	.575	.940
P2- I participate in professional development where I learn to improve my instructional practices.	107.36	329.844	.561	.470	.940
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	108.02	320.407	.622	.474	.940
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	107.58	324.823	.670	.632	.939
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	107.44	326.761	.669	.681	.939
C3- I have a clear understanding of the State Standards for my grade/subject.	107.58	332.529	.498	.401	.941
E1- My instruction intentionally addresses the State Standards for my grade/subject.	107.44	328.512	.565	.587	.940
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	107.80	323.141	.601	.524	.940
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	107.76	318.901	.696	.561	.939
P3- I participate in professional development where I learn how to monitor students' progress.	107.87	319.963	.675	.556	.939
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	107.46	324.377	.664	.600	.939
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	107.49	325.192	.647	.596	.939
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	108.78	322.037	.558	.390	.941
E11- I am involved in meetings where data results are discussed.	107.65	322.203	.645	.556	.939
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	107.85	321.473	.658	.577	.939
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	107.56	327.499	.621	.462	.940

C1- I can summarize my schools' shared vision/mission.	107.75	331.734	.475	.332	.941
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	108.01	327.996	.524	.502	.941
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	108.05	322.659	.574	.426	.940
F2- I regularly communicate with families regarding student academic goals/progress.	107.74	323.755	.618	.664	.940
F3- I regularly communicate with families regarding student behavioral goals/progress.	107.74	323.636	.613	.714	.940
F4- I make informed decisions based on feedback from families.	107.78	324.970	.601	.596	.940
F5- I think my school does a good job of including parents as team members in data-based decision making.	108.33	324.817	.539	.410	.941
C5- I have the technology and resources that I need to provide effective instruction.	108.13	332.995	.361	.269	.943
F1- I consider my students' backgrounds when planning instruction.	107.53	329.817	.543	.401	.940
E4- I review universal screening data at least three times a year for every student that I support.	107.84	320.275	.585	.440	.940
C9- I think that the current school initiatives are improving education for students in my school.	107.95	322.420	.648	.523	.939

Table 4. RTI 28 Items- Item-Total Statistics (Alpha for Scale=0.940)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	97.32	433.135	.585	.535	.938
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	97.92	425.419	.628	.646	.938
P2- I participate in professional development where I learn to improve my instructional practices.	97.43	426.403	.649	.586	.938
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	98.00	427.366	.576	.514	.938
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	97.50	426.117	.723	.688	.937
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	97.54	428.331	.606	.628	.938

C3- I have a clear understanding of the State Standards for my grade/subject.	97.68	433.660	.555	.504	.939
E1- My instruction intentionally addresses the State Standards for my grade/subject.	97.61	424.769	.616	.629	.938
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	97.99	418.280	.710	.675	.937
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	97.89	424.440	.665	.568	.937
P3- I participate in professional development where I learn how to monitor students' progress.	97.95	419.915	.718	.601	.937
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	97.53	423.989	.680	.637	.937
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	97.57	428.120	.660	.609	.937
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	98.65	429.600	.518	.404	.939
E11- I am involved in meetings where data results are discussed.	97.45	426.308	.602	.640	.938
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	97.58	437.983	.476	.572	.939
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	97.50	434.243	.557	.423	.939
C1- I can summarize my schools' shared vision/mission.	97.67	438.251	.455	.353	.940
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	97.80	438.092	.454	.577	.940
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	97.98	425.888	.611	.487	.938
F2- I regularly communicate with families regarding student academic goals/progress.	97.91	429.562	.522	.638	.939
F3- I regularly communicate with families regarding student behavioral goals/progress.	97.89	424.761	.622	.716	.938
F4- I make informed decisions based on feedback from families.	97.84	426.535	.585	.638	.938
F5- I think my school does a good job of including parents as team members in data-based decision making.	98.36	436.746	.453	.396	.940
C5- I have the technology and resources that I need to provide effective instruction.	98.19	443.470	.287	.315	.942
F1- I consider my students' backgrounds when planning instruction.	97.39	428.760	.640	.565	.938

E4- I review universal screening data at least three times a year for every student that I support.	97.41	422.855	.621	.602	.938
C9- I think that the current school initiatives are improving education for students in my school.	98.08	431.978	.533	.630	.939

Table 5. Both State 28 Items- Item-Total Statistics (Alpha for Scale=0.945)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	107.34	344.665	.531	.405	.944
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	107.60	339.302	.626	.533	.943
P2- I participate in professional development where I learn to improve my instructional practices.	107.27	341.113	.613	.489	.943
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	107.93	333.461	.636	.475	.943
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	107.44	339.482	.641	.582	.943
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	107.32	340.555	.656	.634	.943
C3- I have a clear understanding of the State Standards for my grade/subject.	107.45	344.448	.533	.440	.944
E1- My instruction intentionally addresses the State Standards for my grade/subject.	107.32	341.712	.574	.581	.943
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	107.68	334.146	.645	.557	.942
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	107.62	333.342	.699	.548	.942
P3- I participate in professional development where I learn how to monitor students' progress.	107.73	332.321	.703	.587	.942
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	107.33	338.075	.663	.624	.942
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	107.36	339.692	.629	.602	.943
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	108.64	334.022	.565	.394	.944
E11- I am involved in meetings where data results are discussed.	107.56	335.148	.634	.525	.943

C2- I receive school-wide academic and behavioral data in usable and understandable formats.	107.91	334.462	.629	.550	.943
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	107.54	339.000	.638	.471	.943
C1- I can summarize my schools' shared vision/mission.	107.65	342.964	.514	.349	.944
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	107.86	339.226	.576	.544	.943
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	107.98	334.634	.596	.427	.943
F2- I regularly communicate with families regarding student academic goals/progress.	107.66	336.219	.613	.705	.943
F3- I regularly communicate with families regarding student behavioral goals/progress.	107.66	335.405	.627	.748	.943
F4- I make informed decisions based on feedback from families.	107.69	336.284	.629	.600	.943
F5- I think my school does a good job of including parents as team members in data-based decision making.	108.18	336.381	.572	.430	.943
C5- I have the technology and resources that I need to provide effective instruction.	108.05	343.511	.407	.306	.945
F1- I consider my students' backgrounds when planning instruction.	107.44	342.474	.567	.405	.943
E4- I review universal screening data at least three times a year for every student that I support.	107.88	333.567	.548	.406	.944
C9- I think that the current school initiatives are improving education for students in my school.	107.85	335.624	.642	.502	.942

Table 6. Midwestern CW 31 Items- Item-Total Statistics (Alpha for Scale=0.945)

School Implementation Scale Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
C1- I can summarize my schools' shared vision/mission.	125.43880	294.369	.539	.382	.944
C2- I receive school-wide academic and behavioral data in usable and understandable formats.	125.70789	285.092	.678	.572	.943
C3- I have a clear understanding of the State Standards for my grade/subject.	125.21668	298.998	.498	.514	.944
C4- I think my school has an effective process in place to identify available resources (e.g., materials, technology, people).	125.59955	291.355	.588	.563	.943
C5- I have the technology and resources that I need to provide effective instruction.	125.80441	295.600	.410	.338	.946
C6- I have the time necessary to analyze student data and problem-solve with my colleagues.	126.32918	286.468	.579	.427	.944
C7- I am involved in action planning school-wide improvements with the other staff and administrators.	125.72931	288.408	.604	.439	.943

C8- I feel that my administrators are committed to implementing evidence-based instructional practices.	125.10451	299.191	.486	.380	.944
C9- I think that the current school initiatives are improving education for students in my school.	125.55417	289.867	.624	.476	.943
E1- My instruction intentionally addresses the State Standards for my grade/subject.	125.02525	300.340	.511	.552	.944
E2- I am able to differentiate instruction according to student needs while addressing the State Standards.	125.30521	297.133	.556	.410	.944
E3- I monitor each of my students' progress towards meeting the State Standards for my grade/subject.	125.23074	292.563	.670	.590	.943
E4- I review universal screening data at least three times a year for every student that I support.	125.75935	284.828	.590	.446	.944
E5- I review formative assessment data for every student that I support.	125.19591	292.526	.657	.594	.943
E6- I evaluate the effectiveness of my instruction based on common formative assessment data.	125.34580	290.173	.704	.614	.942
E7- I adapt the environment, curriculum, and instruction based on my students' academic data.	125.03835	297.743	.614	.599	.944
E8- I adapt the environment, curriculum, and instruction based on my students' behavioral data.	125.14701	296.826	.572	.480	.944
E9- I modify my instructional practices based on students' common formative assessment data.	125.18792	294.173	.647	.618	.943
E10- Based on assessment results, I re-teach information that students have not mastered.	125.05273	297.732	.571	.484	.944
E11- I am involved in meetings where data results are discussed.	125.33014	289.099	.638	.521	.943
E12- When I'm concerned about a student's academic progress, I collaborate with colleagues to identify interventions.	125.03452	296.352	.596	.608	.944
E13- When I'm concerned about a student's behavioral progress, I collaborate with colleagues to identify interventions.	125.02365	298.091	.565	.584	.944
P1- I participate in professional development where I learn how to develop curricular plans that address the State Standards.	125.39693	290.879	.613	.540	.943
P2- I participate in professional development where I learn to improve my instructional practices.	125.03675	297.422	.575	.466	.944
P3- I participate in professional development where I learn how to monitor students' progress.	125.43337	288.491	.683	.600	.942
P4- I receive coaching/mentoring to implement evidence-based instructional practices.	125.63822	289.436	.601	.461	.943
F1- I consider my students' backgrounds when planning instruction.	125.24896	296.523	.571	.388	.944
F2- I regularly communicate with families regarding student academic goals/progress.	125.29850	293.773	.580	.777	.944
F3- I regularly communicate with families regarding student behavioral goals/progress.	125.32822	293.170	.568	.782	.944
F4- I make informed decisions based on feedback from families.	125.42090	292.321	.616	.572	.943
F5- I think my school does a good job of including parents as team members in data-based decision making.	125.88271	288.226	.595	.491	.944