7-1-2016

Multilevel Analysis of a Scale Measuring Educators' Perceptions of Multi-Tiered Systems of Supports Practices

Leslie Marie Marshall
University of South Florida, lesliemarshall07@gmail.com

Multilevel Analysis of a Scale Measuring Educators' Perceptions of Multi-Tiered Systems of Supports Practices

by

Leslie M. Marshall

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Jose M. Castillo, Ph.D.
George M. Batsche, Ed.D.
Roger Boothroyd, Ph.D.
Robert Dedrick, Ph.D.
Eun Sook Kim, Ph.D.

Date of Approval:
June 13, 2016

Keywords: response to intervention, multilevel
confirmatory factor analysis, implementation fidelity, data-based
decision making.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This study aimed to provide evidence of reliability and validity for the 42-item *Perceptions of Practices Survey.* The scale was designed to assess educators' perceptions of the extent to which their schools were implementing multi-tiered system of supports (MTSS) practices. The survey was initially given as part of a larger evaluation project of a 3-year, statewide initiative designed to evaluate MTSS implementation. Elementary educators (Level-1 n = 2,109, Level-2 n = 62) completed the survey in September/October of 2007, September/October of 2008 (Level-1 n = 1,940, Level-2 n = 61), and January/February of 2010 (Level-1 n = 2,058, Level-2 n = 60). Multilevel exploratory and confirmatory factor analysis procedures were used to examine the construct validity and reliability of the instrument. Results supported a correlated four-factor model: *Tiers I & II Problem Solving*, *Tier III Problem Identification*, *Tier III Problem Analysis & Intervention Procedures*, and *Tier III Evaluation of Response to Intervention.* Composite reliability estimates for all factors across the three years approximated or exceeded .84. Additionally, relationships were found between the *Perceptions of Practices Survey* factors and another measure of MTSS implementation, the *Tiers I & II Critical Components Checklist.* Implications for future research regarding the psychometric properties of the survey and for its use in schools are discussed.

CHAPTER I: Introduction

Policy makers have promoted higher student achievement and have strived to make education in the United States more equitable for all students for the past 50 years. The emphasis on improving student outcomes is evident in two major federal education laws. One law, the Elementary and Secondary Education Act of 1965 (ESEA) and its subsequent reauthorizations, emphasized the importance of raising achievement and closing achievement gaps nationwide. The reauthorization of the ESEA in 2001, the No Child Left Behind Act (NCLB, 2002), increased accountability for students and teachers through mandating the use of yearly standardized tests as a measure of how well schools were performing compared to set achievement standards. Moreover, NCLB required the use of data-based decision making and evidence-based practices to ensure accountability for all students. More recently (in 2011), the U.S. Department of Education (US DoE) allowed state educational agencies (SEAs) to request flexibility waivers that would allow states and their local educational agencies (LEAs) to implement state and local reform initiatives that differed from some of the specific requirements set forth by NCLB. However, in order to get the waivers, states were required to continue emphasizing accountability by adopting rigorous student and educator evaluation practices (USDoE, 2012). The trend toward increased flexibility for states continued with the latest ESEA reauthorization, the Every Student Succeeds Act (ESSA, 2016), but a focus on data-based accountability for student outcomes remained in the legislation.

In addition to general education concerns, policy makers also focused on accountability for the outcomes of students with disabilities. The Individuals with Disabilities Education

Improvement Act (IDEIA, 2004) specifically permitted school districts to use "a process that determines if the child responds to scientific, research-based intervention" (§ 1414[b][6]) for the purpose of identifying a student as having a Specific Learning Disability (SLD). Traditionally, evaluations of eligibility for services under the SLD category used a discrepancy model that required a student to have a significant difference between achievement and intellectual ability. Many argued that the discrepancy model was a "wait to fail" model because students were delayed services until their achievement was considerably low enough for the discrepancy to exist (Reschly, 2008). In fact, the inclusion of language in IDEIA emphasizing the use of "scientific, research-based interventions" underscored the role of general education in ensuring that students are consistently exposed to evidence-based practices prior to consideration for special education services.

To best accommodate the needs of all learners, provide services and interventions to students sooner, and to meet the accountability requirements set forth in federal policy (IDEIA, 2004; NCLB, 2002; USDoE, 2011), a Multi-Tiered System of Supports[1] (MTSS) was suggested as a useful framework (Batsche et al., 2005; Prasse, 2014). MTSS is a term used to describe a comprehensive model of schooling that employs data-based decision making to provide differentiated academic and behavioral supports to students (Batsche et al., 2005). An MTSS is typically comprised of three tiers of services that are distributed on a continuum in proportion to student needs. The first tier includes general academic and behavior instruction and support that is differentiated for all students. Tier II resources are more focused and include targeted instruction and intervention supports that are provided to students in need of resources beyond what is provided to all students. Finally, interventions and supports in Tier III are even more specialized and targeted and are given to a very small percentage of students in addition to what is provided in the first two tiers. Across tiers, decisions are made based on data and student progress is monitored more frequently as students receive more intensive services. In essence, MTSS is a model for systematically making decisions about student services using data and proportionally allocating resources to students based on their needs.

Many MTSS models require educators to use the problem-solving process in order to determine students' needs and how to distribute resources efficiently (Batsche et al., 2007; Brown-Chidsey & Steege, 2010). The problem-solving process commonly includes four steps. The first step is Problem Identification during which what students should know and be able to do are operationally defined and the discrepancy between what is occurring and what is expected is measured. The second step is Problem Analysis. During this stage, hypotheses are developed regarding why the problem is occurring and then the hypotheses are validated or invalidated using multiple sources of data. The third step of the process is Intervention Plan Development and Implementation, which involves developing an intervention plan based on evidence-based strategies/interventions that are matched to students' needs identified during Problem Analysis. The final step of the problem-solving process is the evaluation of students' response to intervention. In this phase, educators determine the amount of student progress after implementing the intervention plan with fidelity. Based on students' progress on the targeted skills, the intervention is continued, modified, or terminated. Importantly, the problem-solving process is iterative and steps can be revisited based on how students respond to intervention. Additionally, it is important to note that the process is applied to groups of students as well as individual students receiving services across tiers in an MTSS.

Research has indicated that when an MTSS is implemented with fidelity, schools and districts see improvements in reading and mathematics outcomes, academically related behaviors, general academic performance, and decreases in retention and special education referral and placement rates (Burns, Appleton, & Stehouwer, 2005; Griffiths et al., 2007; Hughes & Dexter, 2011). Data recently collected indicate that efforts to implement MTSS are occurring throughout the nation. In a 2008 report, approximately 90% of special education state department directors from the 50 United States and the District of Columbia indicated that their states were providing statewide trainings related to an MTSS (i.e., overview of MTSS, progress monitoring, use of data-driven decision-making; Hoover, Baca, Wexler-Love, & Saenz, 2008). In 2009, according to Zirkel and Thomas (2010), 12 states required MTSS for SLD identification in their State laws, four states allowed a combined approach (e.g., MTSS and discrepancy formula), and 20 states permitted the use of a third research-based alternative. Furthermore, in a Response to Intervention Adoption Survey of school districts across the United States (n = 1,390) in 2011, 68% of the district administrators who responded indicated that they implemented MTSS compared to 32% in 2008 (Spectrum K12, 2011). Most districts reported that MTSS was used for customization of instruction for all students (62%), early intervention services and supports (88%), and special education identification (66%). Based on these results, it is evident that MTSS is being implemented throughout the U.S. with an aim of improving outcomes for all students.

MTSS Implementation Fidelity

Although policies support the use of an MTSS, the actual implementation of an MTSS is the responsibility of educators including administrators, teachers, student services staff, and other school personnel. Moreover, it is the responsibility of educators to implement an MTSS with fidelity to ensure that the system is effective in meeting students' needs and improving student outcomes. Questions remain, however, regarding the extent to which educators can implement the model with fidelity (Berkeley, Bender, Peaster, & Saunders, 2009; Glover & DiPerna, 2007; Kratochwill, Volpiansky, Clements, & Ball, 2007; Spectrum K12 School Solutions, 2011; Vaughn & Fuchs, 2006). To determine how well educators implement MTSS with fidelity, it is necessary to have an instrument with evidence of reliability and validity to measure MTSS practices. Through measuring and monitoring the fidelity of implementation, educators can better determine if an absence of improvement in student outcomes is related to poor implementation of the program and how implementation of the program can be improved (Carroll et al., 2007).

Measuring MTSS Practices

There are multiple research methods available for examining educators' practices and the implementation of MTSS. A few possible methods include direct observation, review of permanent products, interviews, and surveys. With each of these research methods, there are various advantages and disadvantages that must be considered (Jackson, 2009).

The first method, observation of practices, is considered the least biased method that provides the most accurate description of behaviors (Noell & Gansle, 2006; Roach, Lawton, & Elliott, 2014). However, observations are vulnerable to observer bias that can be related to inadequate understanding of how to observe the practices. Additionally, the presence of the observer can influence the behavior of the person being observed because s/he may behave differently when being observed. Direct observation does not always result in an adequate sample as well due to the amount of resources required to schedule and conduct observations (i.e., time, personnel, travel costs). Permanent product reviews also are considered one of the less biased fidelity measurement methods (Noell & Gansle, 2006; Roach, Lawton, & Elliott, 2014). However, due to a lack of consensus among researchers and educators regarding the critical components of an MTSS that need to be measured to adequately capture implementation (e.g., specific instruction and intervention strategies, data sources, steps of the problem-solving process), ample documentation or permanent products are not always available for review (Jackson, 2009). Additionally, permanent products may not be readily available due to inadequate record keeping or a lack of understanding of how to record some elements of MTSS. Another possible research method is interviews. Although interviews are thought to capture multiple dimensions of someone's practices, they typically require more resources than are feasible to conduct enough interviews to get a representative sample of practices. Interviews also are highly subject to interview bias (Desimone, 2009).

Surveys are another potential research method that is considered practical for gathering large amounts of representative data (Desimone, 2009). Although surveys that rely on educators to self-report their practices often are upwardly biased and can overestimate educators' practices (Noell & Gansle, 2006; Roach, Lawton, & Elliott, 2014), they can be administered efficiently to large numbers of educators without as much demand on personnel. Surveys can be useful for measuring perceptions of MTSS practices because educators are able to rate their practices across multiple settings anonymously. Therefore, their reports of practices can be generalized across settings and are not influenced by an observer or an interviewer (Desimone, 2009). In fact, Desimone (2009) found that confidential and anonymous data from robust and properly administered observation, interviews, and surveys for behavior-based constructs (e.g., professional development activities, classroom instruction) can provide similar information.

Empirically validated measures of educators' perceptions of MTSS practices are not widely available in the literature. Although self-report instrumentation that measures MTSS implementation as reported by school leadership teams is available (e.g., *Benchmarks of Quality*

[*BoQ;* Kincaid, Childs, & George, 2010]*, Effective Behavior Support Survey* [*EBS*; Sugai, Horner, & Todd, 2003], *Self-Assessment of Problem-Solving Implementation [SAPSI*; Castillo et al., 2010]*, Self-Assessment Tool* [*SET*; Sugai, Lewis-Palmer, Todd, & Horner, 2001]*, Team Implementation Checklist* [*TIC*; Sugai, Horner, Lewis-Palmer, Rossetto, 2012]), little instrumentation exists that can be administered to school staff across a building. Because multiple school personnel implement various components of an MTSS, it is important that the practices of various stakeholders are captured rather than solely capturing the perceptions of a small sample of stakeholders that may have limited perspectives regarding what is being implemented. One instrument known to measure staff perceptions is the *Perceptions of Practices Survey.*

The Perceptions of Practices Survey originally was developed as part of a three-year state-level evaluation of MTSS implementation (Castillo, Hines, Batsche, & Curtis, 2011). The survey was designed to assess educators' perceptions of the extent to which their schools are implementing MTSS practices. Data were to be aggregated to the school-level to inform activities related to facilitating staff consensus regarding implementation as well as facilitating implementation with fidelity (Castillo, et al., 2010). The *Perceptions of Practices Survey* contained 42 items that aimed to measure school staff's perceptions about practices in one or more of the following domains: data-based decision-making, tiered service delivery, the problem-solving process, and special education eligibility determination, as they pertain to academics and behavior. The scale used was a 5-point scale (*1 = Never Occurs to 5 = Always Occurs*). Items for the *Perceptions of Practices Survey* were developed through a review of relevant literature, presentations, instruments, and previous program evaluation projects. The survey was sent to an Educator Expert Validation Panel (EEVP) that reviewed the items for clarity and quality. Based on the EEVP members' feedback, a few items were revised to provide clarification of terminology. Preliminary construct validity evidence for the survey was established through single-level exploratory factor analysis (EFA). Two factors that accounted for 75% of the common variance in perceived practices were established: *Perceptions of RtI Practices Applied to Academic Content* and *Perceptions of RtI Practices Applied to Behavior Content.* Additionally, internal consistency reliability estimates were high for each factor (α = .97 & .96 respectively; Castillo et al., 2010). Although the single-level EFA provided some preliminary evidence for the validity of the tool, further examination of the measure's construct and criterion validity as well as reliability is needed.

Validity and Reliability

Validity is considered one of the most fundamental aspects of developing and evaluating measures. It is defined as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p.11). Multiple forms of validity evidence exist including evidence regarding cognitive processes, internal structure, relationships with conceptually related constructs, relationships with criteria, evidence based on consequences of tests, and content-oriented evidence (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Depending on the intended purpose of the tool being developed and evaluated, different methods are needed to provide evidence of validity.

One source of evidence for validity is construct validity, which involves the examination of the degree to which the tool measures the intended constructs based in theory (Cronbach & Meehl, 1955). Construct validity is typically examined through exploratory or confirmatory factor analysis (EFA or CFA) or cluster analysis. These analyses are employed to determine the underlying factor structure present in the observed variables to better understand the shared variance among the measured variables that are hypothesized to characterize the construct (Thompson, 2004). Traditionally, psychometric studies examining construct validity have inspected the factor structure of an instrument at a single-level. However, conducting psychometric analyses at a single-level with data from educational settings often ignores the potential effects of the school-level variables on individual educators (nested data; Dedrick & Greenbaum, 2011). When nested data are present (e.g.,, teachers nested within schools), multilevel approaches for examining validity evidence may provide more accurate and less biased estimates.

To examine the validity and reliability at multiple levels (i.e., educator and school levels), researchers have suggested using a multilevel EFA or CFA (MCFA; Dedrick & Greenbaum, 2011; D'Haenens, Van Damme, & Onghena, 2012). MEFA and MCFA can account for variance in observed scores at both the educator- and school-levels. More specifically, multilevel factor analysis procedures include within and between latent factors and within and between factor loadings that can assess construct validity at the teacher and school levels. Reliability of latent factors also can be conducted across levels. Examining reliability using a single-level approach (e.g., Cronbach's alpha) for nested data can confound the within-group variance and between-group variance and result in biased reliability estimates because it violates the assumption of independent residuals. MCFA, however, allows for estimates of reliability to be derived at the school-level in addition to the educator-level (Dedrick & Greenbaum, 2011).

In addition to examining construct validity, it is useful to investigate criterion validity, which includes concurrent validity. Concurrent validity involves the degree to which a measure correlates with other measures of the same construct that are assessed at the same time (Mislevy & Rupp, 2010). Researchers have developed other methods for evaluating MTSS implementation fidelity that can be used to investigate the concurrent validity of the *Perceptions of Practices Survey*. One such measure is the *Tiers I and II Critical Components Checklist* (Castillo et al., 2010). The *Checklist* involves permanent product review methodology in which trained reviewers use a standard rubric to evaluate evidence of MTSS practices in documents from data meetings focused on student performance. Given concerns with self-report measures of fidelity raised in the literature, research is needed that relates self-report tools such as the *Perceptions of Practices Survey* to other methods of measuring fidelity such as the *Checklist*.

Purpose

The present study was a secondary analysis of the factor structure and reliability of the *Perceptions of Practices Survey* using data from educators nested within elementary schools in a southeastern state. The *Perceptions of Practices Survey* was intended to examine perceptions of MTSS practices at the school-level through collection of data on the perceptions of individual educators. Because the factor structure of an instrument can differ across units of analysis (Dedrick & Greenbaum, 2011), the factor structure of the tool was examined at both the educator- and school-levels. The extent to which the tool related to another measure of fidelity (i.e., concurrent validity) also was investigated. The research questions addressed through this study were:

1. Is the factor structure (i.e., number of factors and factor loadings) underlying the *Perceptions of Practices Survey* at the educator-level similar to or different from the structure at the school-level?

2. Are the factor loadings underlying the *Perceptions of Practices Survey* significantly different between and within schools?
3. To what extent are the factor structure, factor loadings, and correlations of factors different across multiple time points for the *Perceptions of Practices Survey*?
4. What is the reliability of the scores from the *Perceptions of Practices Survey* at the educator and school levels across multiple time points?
5. To what extent are the factor scores derived from the *Perceptions of Practices Survey* related to the *Tiers I and II Critical Components Checklist*?

Endnotes

[1] Multi-Tiered System of Supports (MTSS) is the most recent term that has replaced the previous term Response to Intervention (RtI), thus MTSS will be used throughout the document in lieu of RtI to reflect the updated terminology.

CHAPTER II: Review of the Literature

Relevant research on a Multi-Tiered System of Supports (MTSS) and educators' perceptions of practices relative to MTSS will be reviewed in this chapter. First, the context for the shift to an MTSS model of service delivery in the United States will be highlighted. Next, a description of the MTSS model of service delivery, the student and systemic outcomes related to an MTSS, and current literature about the fidelity of implementation of MTSS will be reviewed. Finally, the current research on the measurement of practices relative to MTSS will be reviewed and a need for more reliable and valid measures of fidelity of MTSS implementation will be discussed.

Context for the Shift to an MTSS Model

The educational system in the United States is continuously evolving with an increasing focus on improving outcomes for all students. This increasing focus was most evident when the *No Child Left Behind Act of 2001 (NCLB)* was enacted in an attempt to increase accountability for schools and improve education for all students including general and special education students. More specifically, NCLB required the use of standardized assessments to measure student performance (annually for grades 3-8 and once in high school) and evaluate schools. NCLB also required that schools, districts, and states publicly report aggregate test results and disaggregated results by student demographic subgroups, including low-income students, students with disabilities, English language learners, and racial groups. Additionally, as part of NCLB's aim to improve schools, the law required that educational programs and practices be based on scientific research providing evidence of the effectiveness of the educational programs.

In 2004, IDEA was reauthorized and renamed the *Individuals with Disabilities Education Improvement Act (IDEIA)*, which reformed the traditional method for identifying students with disabilities in a manner consistent with the accountability focus of NCLB (2002). Prior to the reauthorization of IDEIA, schools determined learning disabilities based on a "significant discrepancy" between a child's IQ and achievement scores. This model was highly criticized as a "wait to fail" model because children typically struggled two to three years before they exhibited a "significant discrepancy" and were given the services they needed (Lyon et al., 2001). Instead of the traditional "significant discrepancy" formula that was used for identifying learning disabilities based on IQ tests, IDEIA allowed school districts to adopt alternative models such as the Response to Intervention (RTI) model (IDEIA, 2004). The RTI model is a multi-tiered approach that allows schools to implement interventions at increasing levels of intensity to support struggling learners and to identify students with disabilities based on a failure to respond to high-quality, evidence-based interventions that are implemented with fidelity. In addition to recommending the use of RTI, IDEIA also promoted the implementation of early evidence-based intervention services to prevent students from falling too far behind expectations (IDEIA, 2004).

NCLB (2002) and IDEIA (2004) highlighted the importance of high quality, evidence-based instruction and interventions, and held schools accountable for *all* students' progress in meeting rigorous standards. More recently, the U.S. Department of Education (DoE; 2010) published its blueprint for the reauthorization of NCLB. In this blueprint, the U.S. DoE outlined priorities that included raising standards and accountability for all students, improving assessments to measure student growth in relation to college and career readiness, providing an all-encompassing education (e.g., history, language arts, fine arts, technology, financial literacy, mathematics) using evidence-based practices, meeting the needs of diverse learners (e.g., English

Language Learners, students with disabilities, neglected/delinquent students, homeless students), and eliminating the achievement gap. Moreover, the U.S. DoE called for comprehensive student services and the use of data to improve students' safety, health, and well-being.

Despite the vision for the NCLB reauthorization set forth in the U.S. DoE's (2010) blueprint, Congress did not act to reauthorize NCLB (2002) for several years. As a result, the U.S. DoE provided waivers for flexibility from some requirements of NCLB to 43 states, the District of Columbia (D.C.), and Puerto Rico. With these waivers, State Educational Agencies (SEAs) were able to develop and implement "rigorous and comprehensive" plans to improve education for all students (U.S. DoE, 2012). To obtain a waiver, states had to clearly outline their plans including learning standards, student assessments, and school and educator accountability systems. The waivers explicitly required that states adopt rigorous statewide standards and assessments that measure student achievement and growth, an accountability system that identifies high- and low-performing schools and supports the consistently low-performing schools, and educator evaluations that focus on improving instruction. This trend toward state flexibility continued with the recent reauthorization of NCLB, the Every Student Succeeds Act (ESSA, 2016). Although greater flexibility was provided to states, data-based accountability for student performance remained a key component of the legislation.

Overview of the MTSS Model

The MTSS Model incorporates many of the critical components set forth by national legislation (e.g., NCLB, IDEIA) and policy including scientifically-based instruction and interventions matched to students' needs. An MTSS is a multi-tiered approach for identifying students with learning and behavior needs as early as possible through ongoing student assessment (e.g., universal screening and progress monitoring) and providing support for these students through levels of differentiated interventions and supports (Fletcher & Vaughn, 2009; Stoiber, 2014).

The goal of an MTSS is to support all students in meeting their post-secondary goals by distributing resources appropriately to all students based on their levels of need. To accomplish this goal, need-driven decision-making occurs at various levels of the educational system (i.e., district, school, classroom). There are three primary components of an MTSS including a problem-solving framework, a tiered system of supports, and a data collection and assessment system to inform decision-making at each tier (Batsche et al., 2005; Fletcher & Vaughn, 2009; Stoiber, 2014). Notably, different MTSS models exist such as RTI and Positive Behavior Supports (PBS), however, these models have similar components mentioned previously. Currently, the term MTSS is used as a more encompassing term that describes organized tiers of resources using data-based decision making and includes RTI (academics focus) and PBS (behavior/social-emotional focus).

The problem-solving framework is used at multiple levels of the system such as the state-, district-, school-, classroom-, and individual student-level. A commonly used problem-solving framework in an MTSS includes the following four steps: (1) Problem Identification (i.e., well-defined problem that includes the discrepancy between what is occurring and what is expected); (2) Problem Analysis (i.e., hypotheses regarding why the problem is occurring that are either confirmed or rejected based on data); (3) Intervention Plan Development and Implementation (i.e., a plan is developed and implemented that includes evidence-based strategies that are matched to needs determined by validated hypotheses); and (4) Evaluation of the response to intervention (i.e., a decision regarding how much progress was made based on frequently collected data).

Within an MTSS, the problem-solving framework is used to distribute resources and services among multiple tiers. Although different MTSS models exist, most models include three tiers of services (Batsche, Curtis, Dorman, Castillo, & Porter, 2007; Burns & VanDerHeyden, 2006; Fletcher & Vaughn, 2009; Stoiber, 2014). Tier I includes scientifically-based instruction and supports (academic, behavioral/social-emotional) aligned with district curriculum and state standards that are provided to *all* students. Tier II is comprised of "more intensive" instruction and interventions (e.g., more time, narrower focus of instruction) that some students receive in addition to Tier I instruction and supports. Tier II services are given to students in need of additional support to meet performance expectations and can be given in various settings by various professionals. Tier III includes the "most intensive," more individualized services that are provided to few students (e.g., approximately 5% of the students in a school). Instruction and support in Tier III are more time intensive and have a narrower focus than services offered in Tier II to help reduce significant barriers for students so that they are better able to meet expectations.

An integrated data collection and assessment system is crucial for decision-making across tiers in an MTSS (Batsche et al., 2005; Burns & VanDerHeyden, 2006; Fletcher & Vaughn, 2009; Stoiber, 2014). To determine which students are not meeting expectations and are in need of additional supports, educators must engage in ongoing data collection and analysis. An integrated data collection and assessment system provides an efficient method for determining whether or not instruction and supports are effective at Tiers I, II, and III. Within an MTSS, educators use instructionally relevant assessments with evidence of reliability and validity including screening, diagnostic, progress monitoring, formative, and summative measures. At Tier I, educators use universal screening assessments that identify which students are not meeting expectations. At Tiers II and III, various assessments are given more frequently to monitor students' progress. Through using these assessments, educators are able to determine students' responsiveness to instruction/intervention and make meaningful decisions (Burns & VanDerHeyden, 2006a; Fletcher & Vaughn, 2009; Florida's MTSS, 2014; Stoiber, 2014).

Student and Systemic Outcomes in the Traditional Model of Education

Despite policy reforms and the calls for MTSS to be implemented to meet the needs of students, a substantial number of students in the U.S. are struggling to meet basic proficiency (i.e., partial mastery of fundamental skills; Kena et al., 2014). Although scaled scores have gradually increased for 4th and 8th graders in reading, 32% of 4th grade students and 22% of 8th grade students were not meeting basic proficiency in 2013. Although more 4th and 8th grade students were meeting basic proficiency in mathematics as compared to reading, a number of students still were not meeting basic proficiency in math (17% of 4th graders and 26% of 8th graders). Additionally, despite increased reading and mathematics scores for all racial/ethnic groups, there still remain large gaps among demographic groups (see Table 1 for more details; Kena et al., 2014).

Educators and policymakers also paid special attention to services for children and youth with disabilities (Kena et al., 2014). From 1990 to 1995, the number of students (ages 3-21) who received special education services increased from 4.7 million to 6.7 million (Kena et al., 2014). In 2002, the President's Commission on Excellence in Special Education (PCESE) reported that almost half of the six million students being served through special education were identified as having a "specific learning disability." Moreover, from 1976 to 2002, the number of students identified as having a specific learning disability increased more than 300%. Additionally, the traditional discrepancy model of identification for special education resulted in the over-identification of racially and ethnically diverse students, males, English Language Learners, and students from low income families (Donovan & Cross, 2002; Heller, Holtzman, & Messick, 1982). These data seem to indicate that the traditional service delivery model was not sufficient in helping all students achieve academic proficiency, including those students from disadvantaged backgrounds. Despite access to special education services through previous iterations of IDEIA (2004), many researchers called for changes in special education that were included in the 2004 reauthorization of IDEIA (Heller, Holtzman, & Messick, 1982; Hosp & Reschly, 2003; PCESE, 2002; Vaughn & Fuchs, 2003).

Table 1

Average National Assessment of Educational Progress (NAEP) Reading and Mathematics Scale Scores of 4^{th}- and 8^{th}-grade Public School Students in 2013 by Race/Ethnicity

	Asian	White	Hispanic	Black
4^{th}-grade Students Reading Scale Score[1]	237	231	207	205
8^{th}-grade Students Reading Scale Score	280	275	255	250
4^{th}-grade Students Mathematics Scale Score	260	250	230	224
8^{th}-grade Students Mathematics Scale Score	308	293	271	263

[1]Scale scores range from 0 to 500.
Note. Data derived from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2013 Mathematics and Reading Assessments, retrieved March, 2014 from the Main NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

With the authorization of NCLB, the President's Commission on Excellence in Special Education was formed to develop recommendations to improve the education of children with disabilities. In the Commission's (2002) report, "A New Era: Revitalizing Special Education for Children and Their Families," it was recommended that IDEA focus on student outcomes, emphasize early identification and intervention using scientifically-based methods to prevent student failure using models such as RTI, and consider all children general education students first and use general education funds for these students. After the policy reforms that occurred in IDEA, in 2012, the number of students receiving special education services declined to approximately 6.4 million. Of those 6.4 million students, the most common disabilities served included “specific learning disabilities” (36%), “speech or language impairments” (21%), and “other health impairments” (12%; Kena et al., 2014). Thus, after the changes in national education policy, the rates of students identified with a specific learning disability declined from approximately 50% of students to about 36% of students in special education (Kena et al., 2014; PCESE, 2002).

Student and Systemic Outcomes in an MTSS Model

The limited research studies that have examined the impacts of various MTSS models have shown promising results for a range of student and systemic outcomes. Burns, Appleton, and Stehouwer (2005) conducted a meta-analytic review of the effectiveness of MTSS models implemented for research purposes as well as field-based models of MTSS. In their meta-analysis, they examined the effectiveness of an MTSS model related to student and systemic outcomes through a review of studies from four large-scale district or state MTSS implementation initiatives (i.e., Heartland Agency Model, Ohio Intervention Based Assessment Model, Pennsylvania Instructional Support Team Model, Minneapolis Public School’s Problem-Solving Model). Burns et al. (2005) reviewed 21 MTSS implementation studies and found positive results for both research-implemented MTSS models and field-based models. High, unbiased estimates of effect (UEE; a weighted estimator of effect that employs effect size and the sample size for each study) were found for both existing field-based RtI models (1.42) and research-implemented models (.92). Furthermore, positive results were found for both student and systemic outcomes. The authors found higher overall UEEs for systemic outcomes (1.54) than for student outcomes (1.02), yet both UEEs were greater than 1.00, indicating a large effect size for both systemic and student outcomes. However, there were differences in student and systemic outcomes between the two groups. Specifically, although existing field-based models resulted in larger UEEs for systemic outcomes (1.80) than for student outcomes (.94), research-implemented models, on the other hand, had larger UEEs for student outcomes (1.14) than for systemic outcomes (.47). The researchers mentioned that the differences between the two models could be due to the length of program implementation. In this case, it is likely that teams implementing field-based models had more time to refine their implementation models because of the longer period of implementation, which could have resulted in increased systemic outcomes. Additionally, it is possible that the policies and procedures implemented by the school districts and states impacted the systemic outcomes. Rates of special education referrals and placements also were reviewed. Results indicated that 1.68% of the student population was placed into special education, as opposed to previous estimates that about 5% of the student population was identified with a learning disability (Lerner, 2002).

Griffiths, Parson, Burns, VanDerHeyden, and Tilly (2007) also conducted a review of the literature on the effectiveness of MTSS implementation. Through their review, the authors found that the school-based teams served more students and special education rates remained constant, or in some cases, decreased. Additionally, a larger percentage of students referred for evaluation qualified for services in comparison to previous years when the traditional model was employed. Griffiths et al. (2007) also found that some studies reported a decline in grade retention rates. Finally, many of the teachers in the studies reported a better ability to develop quality Individualized Education Plans (IEPs) and interventions based on the data that were collected throughout the process.

Hughes and Dexter (2011) also reviewed field studies examining the effectiveness of MTSS. Included in their review were 13 published studies that included a problem-solving component (*n* = 7), a standard protocol form of MTSS (i.e., preselected interventions that are used when a student fails to respond; *n* = 5), or a combination of both (*n* = 1). Overall, the authors found that the studies that examined the impact of MTSS on students' academic outcomes reported a positive impact on early reading and mathematics skills. The authors also found that special education referral rates either remained constant, or showed slight declines.

Researchers also have found that schools implementing MTSS with effective core curricula had positive improvements in reading test scores for all students (high and low achievers) and the gap between the performance of high and low achievers was reduced throughout the year (Mellard, Frey, & Woods, 2012). Additionally, researchers have shown reduced retention rates in elementary school after two years of MTSS implementation (Murray, Woodruff, & Vaughn, 2010). Reductions in special education referral rates and more accurate special education evaluations (i.e., increased likelihood that a student referred would be found eligible) also have been reported (VanDerHeyden, Witt, & Gilbertson, 2007). Finally, researchers found that with the implementation of an MTSS model, children were identified with learning disabilities earlier allowing for them to receive targeted interventions sooner (Gettinger & Stoiber, 2008; Torgeson, 2009).

Fidelity of MTSS Implementation

Research indicates that when MTSS is implemented with fidelity, schools and districts see improvements in reading and mathematics outcomes, academically related behaviors, general academic performance, and decreases in retention and referral and placement rates (Burns et al., 2005; Griffiths et al., 2007; Hughes & Dexter, 2011). Recent data indicate that efforts to implement MTSS are occurring throughout the United States. In 2008, approximately 90% of special education state department directors from the 50 United States and the District of Columbia indicated that their states were providing statewide trainings related to an MTSS (i.e., overview of MTSS, progress monitoring, use of data-driven decision-making; Hoover, Baca, Wexler-Love, Saenz, 2008). According to Zirkel and Thomas (2010), in 2009, 12 states required MTSS for SLD identification in their laws, four states allowed a combined approach (e.g., MTSS and discrepancy formula), and 20 states permitted the use of a third research-based alternative. Through a Response to Intervention Adoption Survey of school districts across the United States (n = 1,390) in 2011, 68% of the district administrators who responded indicated that they implemented MTSS compared to 32% in 2008 (Spectrum K12, 2011). Most districts reported that MTSS was used for customization of instruction for all students (62%), early intervention services and supports (88%), and special education identification (66%). Based on these results, it is evident that MTSS is being implemented throughout the U.S. with an aim of improving outcomes for all students.

Although implementation of an MTSS has resulted in positive outcomes, some school districts and researchers have been concerned that MTSS is not always resulting in improved outcomes for students (O'Connor & Freeman, 2012). This suggested lack of improvement in student outcomes may be due to MTSS not being implemented with fidelity, or in other words, the critical components of MTSS have not been implemented in the manner that they are intended to be implemented (Berkeley, Bender, Peaster, & Saunders, 2009; Glover & DiPerna, 2007; Kratochwill, Volpiansky, Clements, & Ball, 2007; Spectrum K12 School Solutions, 2011; Vaughn & Fuchs, 2006). To determine if an intervention or program results in anticipated outcomes and to determine the particular parts of a program that result in those outcomes, it is necessary to evaluate the fidelity of implementation of the program. If fidelity of implementation is not measured, then it is difficult to determine whether the lack of impact is due to components of the program itself or poor implementation of the program (Carroll et al., 2007).

The fidelity of MTSS implementation is arguably one of the most important components of implementation efforts, yet fidelity has received little attention in comparison to other elements of implementation (Gansle & Noell, 2007). Although there is a lack of consensus regarding a definition of fidelity of MTSS implementation, fidelity commonly captures a range of complex activities that are required to implement a model of MTSS (Keller-Margulis, 2012). The fidelity of implementation of MTSS components has received some attention in the literature (e.g., Duhon, Mesner, Gregerson, & Witt, 2009; Fletcher, Francis, Morris, & Lyon, 2005; Glover & DiPerna, 2007; Kelleher, Riley-Tillman, & Power, 2008; Noell et al., 2005). However, most of the literature focuses on the fidelity of intervention implementation or fidelity of individual MTSS components rather than the fidelity of implementation of multiple components of an MTSS model (e.g., problem-solving, tiered levels of supports, assessments, decision making procedures) as a whole.

In order to assess the effectiveness of the implementation of an MTSS model, Keller-Margulis (2012) suggested that the general domains and critical components that exist among the multiple tiers of an MTSS model must be assessed including "(1) assessment practices, (2) instruction and intervention delivery, and (3) procedural decision making" (p. 345). VanDerHeyden, Witt, and Gilbertson (2007) were the first to attempt to examine the fidelity of an MTSS system rather than monitoring the fidelity of intervention components alone. In their study, VanDerHeyden, Witt, and Gilbertson (2007) examined the fidelity of implementation of the System to Enhance Educational Performance (STEEP), which is a program that incorporates components of an MTSS model. In this study, fidelity of assessment techniques (e.g., universal screening) and decision-making (e.g., evaluating students' response to intervention) procedures were conducted through observations by trained observers using checklists and through calculating the percent agreement between judgments of responsiveness to intervention. Although it is recognized that multiple domains are important components of MTSS models that should be monitored for fidelity, there is a dearth of research indicating efficient and appropriate methods for measuring the fidelity of implementation of the domains altogether.

Measuring Fidelity

Measuring fidelity is important to gain information regarding quality of program implementation, examine the foundational theories underlying the program, and to provide feedback for continuous improvement (James Bell and Associates, 2009). To examine the fidelity of implementation of MTSS, it is necessary to have an accurate and statistically adequate measure of practices. Mowbray et al. (2006) suggested that there are three critical steps to adequately measure the fidelity of practices: identify indicators or critical components of the program, gather data to measure the indicators, and examine the reliability and validity of the indicators. Once the critical components are identified, multiple sources of data collection should be considered for gathering the necessary data.

Multiple methods of measuring MTSS practices have been utilized, including observations, interviews, reviews of permanent products, and self-report measures. Observations of practices are generally considered to be the most "accurate" method of measurement for behaviors (Noell & Gansle, 2006; Roach, Lawton, & Elliott, 2014); however, limitations of observation data exist. For instance, observer bias and effects of an outside observer on educators' behaviors can influence observation data. Additionally, observation data can be limited due to the high cost of conducting multiple observations and limited availability of trained observers. Moreover, when examining MTSS practices specifically, it is difficult to observe all practices and know whether or not the practices consistently occur because the practices can span a lengthy period of time (e.g., working through problem solving process as a team and monitoring intervention plans can take six weeks or more). Due to the aforementioned barriers to observations, alternative methods are needed that are more feasible yet valid (Berkel et al., 2011).

Interviews are another common method used to measure educators' practices. Interviews are useful because the researcher can gather in depth qualitative data regarding educator practices. However, interview data can be limited in that they are subject to interviewer bias and the interviewee's potential desire to appear "good" (Desimone, 2009). Additionally, interviews require extensive time and resources and provide data from only a few perspectives. Another potential method for measuring educators' MTSS practices is to review permanent products. A review of permanent products can provide a sample of MTSS practices across time and settings, but permanent products are limited in that they can provide false data (e.g., not completed correctly, over-reporting practices) or they may fail to characterize all practices related to MTSS (Noell & Gansle, 2006; Roach, Lawton, & Elliott, 2014). Reviewing permanent products also is not the most efficient method of data collection because it requires additional time for staff to collect the permanent products, review the permanent products, and rate the level of implementation based on the products. Additionally, research indicates that there can be considerable differences between permanent product reviews and direct observations indicating that permanent product reviews may not provide accurate data regarding implementation (Sanetti & Collier-Meek, 2014).

A final method commonly used to gather data regarding educators' practices is through self-report measures. Despite the fact that some researchers consider self-report measures biased measures of practices because reporters may answer questions in a socially desirable manner (Noell & Gansle, 2006), others consider self-report measures desirable for collecting data on large samples because they require less resources, provide quantitative measures of practices, and have potential to provide accurate information regarding practices (Desimone, 2009). Desimone (2009) found that confidential and anonymous data from robust and properly administered observation, interviews, and surveys for behavior-based constructs (e.g., professional development activities, classroom instruction) can provide similar information. Desimone (2009) further noted, "social desirability bias can occur in any form of data collection" (p. 189). Thus, self-report measures may be adequate measures of educators' practices related to MTSS.

In sum, there are multiple methods for examining the fidelity of implementation of a program. Each method appears to have strengths and weaknesses related to accuracy and efficiency. Of all the methods previously mentioned, self-report is a desirable method of data collection for large samples due to the less obtrusive nature of the method and the lesser amount of resources required, which can result in the ability to gather data more frequently. Furthermore, self-report data can be quantified making it easier to more objectively measure and examine changes in implementation.

Measuring Fidelity of MTSS Practices

Multiple measures have been created to assess the fidelity of implementation of various tiered models. One example is process checklists that measure whether or not the foundational features of the tiered service delivery model are in place. For instance, the *Team Implementation Checklist (TIC)* can be completed once a quarter by the school's Positive Behavior Intervention Supports (PBIS) team to rate activities related to PBIS implementation as "achieved," "in progress," or "not yet started" (Sugai, Horner, Lewis-Palmer, & Rossetto, 2012). Measures of quality and quantity also have been developed that assess the degree to which the model has been implemented school-wide. Examples of these measures for PBIS are the *Self-Assessment Tool* (*SET*; Sugai, Lewis-Palmer, Todd, & Horner, 2001), the *Benchmarks of Quality* (*BoQ*; Kincaid, Childs, & George, 2010), and the *Effective Behavior Support Survey* (*EBS*; Sugai, Horner, & Todd, 2003). The *SET* is a 28-item observation and interview instrument that is completed by a trained coach/researcher. For the *BoQ,* PBIS team members rate each of the 53-items as "in place," "needs improvement," or "not in place." The *EBS* is a tool that is used to rate the current status and priorities for PBIS and can be completed by all school staff.

Although there are multiple tools that have been created to assess the implementation of PBIS specifically, few empirically validated measures exist that evaluate the school-wide implementation of an integrated academic and behavior MTSS model in a cost-effective, minimally intrusive manner. The *Tier I and II Observation Checklist* (Castillo et al., 2010) is one observation tool that can be used to examine fidelity of MTSS implementation. The instrument has 20 items that assess roles and responsibilities of stakeholders and components of the problem-solving process that should be present during Tier I and II problem-solving data meetings. During the data meetings, a trained observer marks each item as "present" or "absent."

Tools are also available for trained reviewers to assess the extent to which fidelity is evident in permanent products from data review meetings. These tools include the *Tier I and II Critical Components Checklist* and the *Tier III Critical Components Checklist* (Castillo et al., 2010). For these checklists, a trained reviewer examines the permanent products from either Tier I or II data meetings or Tier III data meetings and rates each of the items (11 items for the *Tier I and II Critical Components Checklist* and 16 for *Tier III Critical Components Checklist*) that address the problem-solving process as *0 = Absent; 1 = Partially Present;* or *2 = Present.*

The *RTI Fidelity of Implementation Rubric* (Center on Response to Intervention, 2014) was also developed in an effort to examine the fidelity of MTSS implementation. The rubric is intended for use by school personnel responsible for monitoring the fidelity of school-wide implementation of an MTSS model, particularly MTSS coordinators or evaluators with MTSS experience. The rubric is organized according to key components of MTSS implementation identified by the National Center on Response to Intervention (NCRTI) including screening, progress monitoring, multi-level prevention system, and data-based decision making. For each factor, there are descriptions for each level of the ratings of 1, 3, and 5 that the rater chooses to rate the level of implementation of the factor based on interviews with a school's MTSS leadership team and/or site visits.

Another available self-report measure that can be used for measuring the fidelity of MTSS implementation is the *Self-Assessment of Problem-Solving Implementation* (Castillo et al., 2010). The *SAPSI* contains 27 items that assess the consensus, infrastructure, and implementation of MTSS practices and procedures. The assessment tool is completed collaboratively by the School-Based Leadership Team who rates each item as *N= Not Started* (the activity occurs less than 25% of the time); *I= In Progress* (the activity occurs approximately

25% to 74% of the time); *A= Achieved* (the activity occurs approximately 75% to 100% of the time); or *M= Maintaining* (the activity was rated as achieved last time and continues to occur approximately 75% to 100% of the time).

The previously mentioned measures that exist capture limited aspects of integrated MTSS models and are typically completed by only a few members of school personnel (i.e., the leadership team) who may have limited perspectives regarding school-wide implementation. Because multiple school personnel implement various components of an MTSS, it is important that the practices of various stakeholders are captured rather than solely capturing the perceptions of a small sample of stakeholders that may have limited perspectives of what is being implemented. To adequately capture the school-wide implementation of an integrated MTSS model, a cost-effective, efficient, and valid method of data collection is needed that multiple stakeholders can complete.

One self-report tool that was developed in an effort to evaluate the school-wide implementation of MTSS was the *School Implementation Scale* (Erickson, Noonan, & Jenson, 2012). The *School Implementation Scale* is a 42-item online survey that requires educators to rate each statement on a scale ranging from "NOT true of me" to "very true of me." Items on the scale relate to school culture, ongoing professional development, evidence-based practices, and family engagement, all of which are thought to be critical components of effective multi-tiered models in school systems (Erickson, Noonan, & Jenson, 2012; Jenson, 2008). Although the *School Implementation Scale* is a useful tool for gathering data regarding multiple components of MTSS, the items on the survey do not specifically address some of the nuances of problem-solving (e.g., clearly defining the problem, developing hypotheses, developing interventions based on confirmed hypotheses) and tiered levels of supports (e.g., school-wide supports, targeted supports) that are foundational to many MTSS models.

Another self-report measure developed to examine the extent to which MTSS practices are implemented school-wide and gather more detailed information regarding problem solving and tiered levels of supports is the *Perceptions of Practices Survey. The Perceptions of Practices Survey* originally was developed as part of a three-year state-level evaluation of MTSS implementation (Castillo, Hines, Batsche, & Curtis, 2011). The survey was designed to assess educators' perceptions of the extent of implementation of MTSS practices within their schools. Data were to be aggregated at the school-level to inform activities related to facilitating staff consensus regarding implementation as well as facilitating implementation with fidelity. The *Perceptions of Practices Survey* contained 42 items using a 5-point scale (*1 = Never Occurs* to *5 = Always Occurs*) that measured school staff's perceptions about practices in one or more of the following domains: data-based decision-making, tiered service delivery, the problem-solving process, and special education eligibility determination, as they pertain to academics and behavior.

Items for the *Perceptions of Practices Survey* were developed through a review of relevant literature, presentations, instruments, and previous program evaluation projects. An Educator Expert Validation Panel (EEVP) reviewed the items for clarity and quality. A few of the items were revised for clarification of terminology based on feedback from the EEVP. Preliminary construct validity evidence for the survey was established through single-level exploratory factor analysis, which established two factors (*Perceptions of RtI Practices Applied to Academic Content,* and *Perceptions of RtI Practices Applied to Behavior Content*) that accounted for 75% of the common variance in perceived practices. Additionally, internal consistency reliability estimates were high for each factor (α = .97 & .96 respectively; Castillo et al., 2010). Although the *Perceptions of Practices Survey* may be a useful tool for measuring MTSS practices, additional evidence for the reliability and validity of the tool is needed before the self-report measure should be used to measure fidelity effectively (Mowbray et al., 2006).

Measuring Reliability and Validity

The validity of an assessment tool is the degree to which the tool measures the constructs it is intended to measure, or the accuracy of the assessment tool (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Validation is a process employed by a test developer to gather evidence to support the inferences that are to be made from test results (Cronbach, 1971). Various forms of evidence for validity exist including evidence regarding cognitive processes, internal structure, relationships with conceptually related constructs, relationships with criteria, evidence based on consequences of tests, and content-oriented evidence (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). Depending on the intended purpose of the tool being developed and evaluated, different methods are needed to provide evidence of validity.

Three common types of validation studies include: content validation (an examination of the measurement tool to determine if it covers a representative sample of content related to the construct to be measured, typically examined by an expert review panel), criterion-related validation (examining the correlation between the assessment tool and existing criterion variable(s) that represent the construct), and construct validation (examination of the degree to which the tool measures the intended constructs based in theory; Cronbach & Meehl, 1955).

In the fidelity literature, there are five commonly used methods for assessing the reliability and validity of fidelity measures (Mowbray et al., 2003). The first method, exploring reliability, involves examining reliability across respondents, calculating the inter-rater agreement through coefficient kappa, intra-class correlations (ICC), percent agreement, or Pearson correlations. Internal consistency reliability estimates (e.g., Cronbach's alpha) also have been used to determine reliability.

The second through fifth approaches examine the validity of a fidelity measure (Mowbray et al., 2003). The second approach involves using exploratory or confirmatory factor analysis (EFA or CFA) or cluster analysis to determine the underlying factor structure present in the observed variables to better understand the shared variance among the measured variables that are hypothesized to characterize the construct (Thompson, 2004). A third approach requires a comparison of fidelity across groups that are expected to be different (e.g., treatment and control groups). Another validation approach commonly used is convergent validity in which two sources of data about the program are compared to determine how well the data align. The final commonly used approach for assessing fidelity is to examine the relationship between measures of fidelity and intended outcomes to determine if fidelity of implementation resulted in desirable outcomes.

Need for Multilevel Analyses

Schools are social systems because they are organized around integral people (e.g., students, teachers, administrators, student services personnel, parents) who reciprocally interact to achieve defined outcomes (Castillo & Curtis, 2014). Moreover, while school buildings contain smaller systems (e.g., grade-level teams, problem-solving teams), they are simultaneously part of the larger system that is the school district. Thus, educators are part of a complex hierarchical system in which educators are nested within schools, which are nested in districts, which are nested within states. Each component of the system influences each other as well as the larger system as a whole. Therefore, the implementation of MTSS in a district is dependent upon the myriad parts of the system and can vary due to the differences found in the various parts of the system (e.g., leadership, available data, policies and procedures, knowledge and beliefs of educators; Castillo & Curtis, 2014).

Because of the hierarchically nested structure of the school system, it is rare that one level of the system (e.g., educators) is not systematically influenced by another level (e.g., schools). Hence, when researching constructs within the educational context, it is important to consider and account for the multiple levels of influence within the system. In essence, multiple levels of influence must be taken into account when they exist or else incomplete or misspecified models will result (Kozlowski & Klein, 2000).

To assess MTSS practices within a school, it is advantageous and efficient to collect survey data at the individual-level. MTSS practices include shared responsibilities among multiple stakeholders in a school (e.g., administrators, data specialists, general education and special education teachers, school psychologists, counselors), all of whom have various roles and responsibilities and consequently varying perspectives of the practices occurring in a school. Accordingly, it is beneficial to gather the perspectives of all stakeholders and aggregate individual-level data to get a more complete view of the MTSS practices in a school.

Typically, researchers have used single-level confirmatory factor analysis (CFA) to determine latent factors when expectations regarding the factor structure exist. With this method, the latent factors only represent individual teachers' perceptions of their school environment rather than the school processes altogether. Although this approach is commonly employed, there are problems that arise when there is hierarchically nested data (i.e., teachers nested in schools nested in districts; Kozlowski & Klein, 2000; O'Connell & McCoach, 2008). First, statistically, this method violates the assumption of independence because teachers are nested within schools that are nested within districts. Thus, the teachers' survey data are likely correlated because they implement common school-wide and district-wide practices. Second, different latent factors may be present at various levels of analysis (e.g., time, teacher, school; Dedrick & Greenbaum, 2011; D'Haenens, Van Damme, & Onghena, 2012; Kozlowski & Klein, 2000; Raudenbush & Bryk, 2001). To overcome these issues and examine the factor structures of a measurement tool at educator- and school-levels, researchers have suggested using a multilevel CFA (MCFA; Dedrick & Greenbaum, 2011; D'Haenens, Van Damme, & Onghena, 2012; O'Connell & McCoach, 2008; Raudenbush & Bryk, 2001). Using MCFA to analyze the statistical adequacy of a survey would yield more accurate factors at the school-level. Moreover, employing an MCFA results in an analysis that includes the within and between latent factors and within and between factor loadings that assess validity at the educator- and school-levels.

It also is important to use multilevel analyses to assess the reliability of latent constructs due to the nested structure of the data. For example, teachers' ratings in one school may be more alike than teachers' ratings in another school. Examining reliability using a single-level approach (e.g., Cronbach's alpha) for nested data can confound the within-group variance and between-group variance and result in biased reliability estimates because it violates the assumption of independent residuals. Through use of the MCFA, an estimate of reliability within- and between-clusters can be derived, thereby allowing reliability estimate to be acquired at each level (i.e., teachers and schools; Dedrick & Greenbaum, 2011).

Conclusion and Rationale

Schools across the nation are widely adopting MTSS models as the primary method of determining services and supports for students. Due to the widespread adoption of integrated MTSS models, there is a need for cost-effective measures that capture school-wide implementation from various stakeholders to make data-based decisions, determine the fidelity of implementation of multiple components of the MTSS model, and evaluate the effectiveness of implementation. Likewise, in this era of accountability, it is necessary to pinpoint why a program succeeds or fails at producing positive outcomes. In order to determine the aspects of the program that impact outcomes, it is crucial to measure the fidelity of implementation of the program. Through monitoring the fidelity of implementation, it can then be determined whether or not a lack of impact was due to poor implementation or aspects of the program itself. Accurate measurement of fidelity requires the development of measures that are reliable and valid. When evaluating the reliability and validity of measures, it is essential that multilevel psychometric analyses be conducted when data are nested. The current study aims to address these needs by examining the reliability and validity of the *Perceptions of Practices Survey* using multilevel psychometric analyses.

CHAPTER III: Method

This study examined data gathered from a larger evaluation project of a 3-year, statewide initiative designed to evaluate MTSS implementation—the Florida Problem-Solving/Response to Intervention (PS/RtI) Project. As part of the comprehensive program evaluation project, evaluators developed several measures of MTSS implementation fidelity for use in schools (see Castillo et al., 2010). Two of the instruments developed as part of the project were used in the current study. The *Perceptions of Practices Survey* was the focus of the current study. The other instrument, the *Tiers I and II Critical Components Checklist* was used as a measure for concurrent validity analysis. What follows is a description of the methods relevant to the proposed study.

Participants

Participants in the study were part of a longitudinal study—conducted by the PS/RtI Project, a grant-funded project approved by the Florida Department of Education (FLDOE)—that initially was intended to evaluate the implementation of MTSS in pilot schools and demonstration districts in the state of Florida (Batsche et al., 2007). PS/RtI Project staff provided trainings on implementation of an MTSS model four to five days per year for three years to School-Based Leadership Teams (SBLTs) in pilot schools. SBLT members included school staff (e.g., administrators, school psychologists, reading and math coaches) that was responsible for facilitating MTSS implementation in their schools. The Project staff and district-based MTSS Coaches also provided technical assistance to SBLT members and pilot school instructional staff to facilitate the implementation of MTSS following trainings. Comparison schools also were included to provide a referent against which to evaluate MTSS implementation. Comparison schools received no training or technical assistance from Project staff. See Castillo et al. (2011) for more information on the larger evaluation study.

Schools. The number of schools that participated varied slightly across the three-year period. Sixty-two (40 pilot schools), 61 (34 pilot), and 60 (34 pilot) elementary schools participated during Years 1, 2, and 3, respectively. Schools were located in eight, seven, and seven demonstration school districts, respectively, and were selected from a competitive application process to participate in the MTSS evaluation project. Diversity in the schools was represented in terms of geography (districts were located across the southern, central, and northern regions of the state) and student populations served (schools were diverse in terms of race, free-reduced lunch status, etc.). During Year 1, for example, school size ranged from 284 to 1,316 students (M = 686, SD = 224). Student characteristics were diverse across the participating schools with approximately 52% (SD = 28) of the students being White, 25% (SD = 27) African American, 12% (SD = 10) Hispanic, 2% (SD = 2) Asian, 0.2% (SD = 0.3) Native American, and 4% (SD = 2) multi-racial. The percentage of students (a) eligible for free-reduced lunch was 52% (SD = 26), (b) who were English language learners was 9% (SD = 12), and (c) identified as having a disability was 16% (SD = 6; Castillo et al., 2015).

Educators. An average of 2,076 educators participated across the schools. On average, there were 32-34 educators per school. Participants in the proposed study were educators working within each of the participating schools. Educators included administrators, teachers (general and special education), student support services personnel (e.g., school psychologists, guidance counselors), and other instructional positions (e.g., intervention specialists, instructional coaches). Non-instructional staff (e.g., school janitors, paraprofessionals) were not included. A summary of the characteristics of the educator participants is provided in Table 2 for September/October of 2007 (n = 2,140), September/October of 2008 (n = 2,001), and January/February of 2010 (n = 2,088). The majority of survey respondents during September/October 2007 and September/October 2008 data collection waves were general education teachers (69-70%). Most participants' highest degree earned was a bachelor's or master's degree (greater than 90%) and participants' years of experience ranged from 1-24 years. Throughout the three years of data collection, the schools remained constant; however, some educator turnover occurred. Therefore, although the demographics of the participants remained similar, some individual survey respondents changed during the three years.

Table 2

Demographic Characteristics of Survey Respondents

	September/October 2007[a] *n* (%)	September/October 2008[b] *n* (%)	January/February 2010[c] *n* (%)
Position			
MTSS Coach	14 (0.87)	4 (0.30)	3 (0.23)
School Counselor	44 (2.74)	32 (2.43)	32 (2.43)
Principal	33 (2.06)	30 (2.28)	24 (1.82)
General Education Teacher	1151 (71.71)	1003 (76.16)	1034 (78.57)
School Psychologist	30 (1.87)	18 (1.37)	18 (1.37)
Assistant Principal	28 (1.74)	12 (0.91)	14 (1.06)
Special Education Teacher	191 (11.90)	143 (10.86)	116 (8.81)
School Social Worker	11 (0.69)	6 (0.46)	5 (0.38)
Other	103 (6.42)	69 (5.24)	70 (5.32)
Highest Degree Earned			
Bachelor's	964 (59.88)	799 (60.58)	820 (62.22)
Master's	584 (36.27)	487 (36.92)	463 (35.13)
Ed.S.	43 (2.67)	22 (1.67)	24 (1.82)
Doctorate	15 (0.93)	6 (0.45)	5 (0.38)
Other	4 (0.25)	5 (0.38)	6 (0.46)
Years of Experience			
<1	79 (4.90)	12 (0.91)	26 (1.97)
1-4	329 (20.40)	218 (16.52)	227 (17.24)
5-9	353 (21.88)	291 (22.05)	320 (24.30)
10-14	264 (16.37)	237 (17.95)	237 (18.00)
15-19	188 (11.66)	177 (13.41)	154 (11.69)
20-24	168 (10.42)	161 (12.20)	155 (11.77)
≥ 25	232 (14.38)	224 (16.97)	198 (15.03)

Note. Ed.S. = Educational Specialist; MTSS = Multi-Tiered System of Supports.

Instrumentation

Perceptions of Practices Survey. The *Perceptions of Practices Survey* was designed to assess educators' perceptions of the extent to which their schools were implementing MTSS practices. The survey contained 21 items that were intended to measure school personnel's perceptions about practices in one or more of the following domains: data-based decision-making, tiered service delivery, the problem-solving process, and special education eligibility determination, as they pertain to academics and behavior. For each item, educators rated the item for (a) academics and (b) behavior. The 5-point scale used ranged from *1 = Never Occurs* to *5 = Always Occurs*. A *Don't Know* option also was available for educators to select when they were unsure of the extent to which the school was engaging in a particular practice.

Project staff first developed items that represented critical MTSS practices based on literature reviews (i.e., journal articles and book chapters), reviews of presentations related to MTSS, instruments, and previous program evaluations. Next, the drafted items were sent to an Educator Expert Validation Panel (EEVP) that consisted of 14 educators (e.g., general and special education teachers, school- and district-level administrators, student support services personnel, content specialists) from a nearby school district who had background knowledge regarding and experience with MTSS. Devellis's (2012) guidelines were used to gather feedback from the EEVP on the representativeness of the items, clarity and quality of the individual items, and suggested modifications to items. Participants were asked to "rate each proposed item as *Good* (item is clearly and accurately written); *Redundant* (there are items with similar content and meaning); *Nonessential* (the content is non-related to any of the domains); *Poorly Written* (item has semantic or grammatical errors); or *Ambiguous* (item has abstract or vague content, or is double-barreled)" (Castillo et al., 2015, p. 9)[2]. For any response other than *Good*, participants were asked to suggest modifications to the item or write, "delete item." Participants also were given the opportunity to suggest additional items that they believed to be important to accurately measure MTSS related practices (Castillo et al., 2015).

The EEVP provided feedback on the representativeness of the items and the clarity and quality of the individual items as well as suggested modifications to items. EEVP members' feedback was analyzed and revisions to the survey were made using a structured process. Items that had 80% member agreement indicating that the item was relevant and well written were kept. Any items with less than 80% agreement were reviewed and discussed among a team of Project staff. Suggested changes to improve the clarity of the item without altering the meaning of the item were made. EEVP members' suggestions were then compared to the revised item to determine if the disagreements had been resolved. Members whose concerns were resolved were then added to the number of members that initially agreed with the item and the percentage of members in agreement was recalculated. As a result of this process, nine items were revised. Revisions included spelling out acronyms (i.e., "CBM" was spelled-out as Curriculum Based Measures), defining terms (i.e., benchmarks), and adding or altering words (e.g., changing "data are collected to" to "data are used to"). See Appendix A for the *Perceptions of Practices Survey*.

Tiers I and II Critical Components Checklist. The *Tiers I and II Critical Components Checklist* was developed as part of the larger 3-year study. The *Checklist* was designed to measure implementation fidelity of the data-based decision-making processes that inform instruction and intervention in MTSS models (Castillo et al., 2011). The instrument contained 15 items that examined the extent to which each of the four stages of problem solving (i.e., Problem Identification, Problem Analysis, Intervention Development & Implementation, and Program Evaluation/RtI) were observed in the permanent products reviewed. The *Checklists* were completed by trained MTSS Coaches who received training and ongoing technical assistance from study personnel on MTSS, permanent product review procedures, and on completion of the *Checklist*. The MTSS Coaches examined permanent products (e.g., charts/graphs, meeting notes, meeting worksheets) from school data meetings in which grade-level student data were analyzed for the purpose of informing Tier I instruction and/or Tier II intervention. MTSS Coaches used a 3-point-scale scoring rubric (*0 = Absent; 1 = Partially Present; 2 = Present*) to evaluate implementation of critical MTSS components using available permanent products. For selected items, Coaches could select *Not Applicable* if a defensible decision was made to not address a specific critical component. Internal consistency estimates (using Cronbach's alpha) for the total scores used in the current study exceeded .80. See Appendix B for the *Tiers I & II Critical Components Checklist.*

Data Collection Procedures

The *Perceptions of Practices Survey* was administered during September/October 2007, September/October 2008, and January/February 2010. SBLT members and other instructional school staff in pilot and comparison schools completed each survey. Project trainers collected survey data from SBLT members at SBLT Trainings. MTSS Coaches collected survey data from the remaining instructional staff. Staff and grade-level meetings were used to administer the surveys whenever possible. Dissemination through staff mailboxes with directions for completing and returning the surveys was used when administration at staff or grade level meetings was not possible. These procedures resulted in return rates of 45% for September/October of 2007, 48% for September/October of 2008, and 52% for January/February of 2010.

MTSS Coaches completed the *Checklists* during September/October 2007, September/October 2008, and January/February 2010. Coaches coordinated with school personnel to identify and collect permanent products (e.g., data printouts and charts, worksheets, forms, action plans, notes) from grade-level data meetings focused on Tier I instruction and/or Tier II intervention. After permanent products were collected, the Coaches completed the *Checklists* as mentioned previously. Inter-rater agreement procedures were conducted to check for accuracy on checklists completed during the middle of each year. Project personnel or another MTSS Coach independently completed *Checklists* on a subset of grade-level products and calculated inter-rater agreement. Because inter-rater agreement procedures occurred only during the middle of the year, there are no estimates for fall and spring; however, average inter-rater agreement estimates across the three years of the macro-level evaluation study exceeded .90.

Trained Graduate Assistants entered and checked the data for accuracy. Graduate assistants randomly selected 10% of the entered data for accuracy checks. When accuracy fell below 95%, all data in the section being reviewed were examined for accuracy. Accuracy exceeded .95 for data entry.

Analyses

A series of statistical analyses were performed in order to answer the following research questions posed in this study:

1. Is the factor structure (i.e., number of factors and factor loadings) underlying the *Perceptions of Practices Survey* at the educator level similar to or different from the structure at the school level?
2. Are the factor loadings underlying the *Perceptions of Practices Survey* significantly different between and within schools?

3. To what extent are the factor structure, factor loadings, and correlations of factors different across multiple time points for the *Perceptions of Practices Survey*?
4. What is the reliability of the scores from the *Perceptions of Practices Survey* at the educator and school levels across multiple time points?
5. To what extent are the factor scores derived from the *Perceptions of Practices Survey* related to the *Tiers I and II Critical Components Checklist*?

Preliminary analyses. Preliminary analyses were conducted using IBM SPSS Statistics, Version 22.0. Data were examined for accuracy by examining the ranges for each variable to make sure the values fell within the expected ranges. Means, standard deviations, and additional descriptive data were calculated for the sample for all variables of interest. Skewness and kurtosis values were calculated to examine the extent to which the data were normally distributed. Item correlations also were reviewed to determine the relationships among items. Finally, correlations were conducted between items with *Don't Know* responses to determine if there was a response pattern among different types of educators.

Intraclass correlations (ICCs) for each of the items also were examined. ICCs close to zero indicate that nearly all variation is at the educator level, whereas ICCs close to 1.00 indicate that nearly all variation is at the school level. ICCs at or above .05 indicate that multi-level factor analysis is an appropriate method (Dedrick & Greenbaum, 2008; Dyer, Hanges, & Hall, 2005).

Multilevel exploratory factor analysis. Multilevel exploratory factor analyses were conducted with the September/October 2007 sample using Mplus Version 7.3 (Muthén & Muthén, 1998-2012) as a first step to addressing *Research Question 1.* This method was used because educators (Level 1) were nested within schools (Level 2). Model fit was evaluated using the χ^2 likelihood ratio statistic, Bentler's (1990) comparative fit index (CFI), the Bayesian Information Criterion (BIC; Schwarz, 1978), sample adjusted BIC (saBIC), Akaike's Information Criterion (AIC; Akaike, 1987), the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), and the standardized root mean square residual (SRMR-within and -between). CFIs greater than or equal to .95 and SRMR and RMSEA values less than or equal to .08 (Hu & Bentler, 1999) were used as the criteria to indicate good levels of fit. The BIC and AIC were used to compare the relative fit of models derived from the MEFA (lower values indicated better relative fit). To determine whether the items contributed to resultant factors, a criterion of .30 (Hair, Tatham, Anderson, & Black, 1998) for the item pattern coefficients (i.e., factor loading) was used. The models resulting from the MEFA were critically evaluated for theoretical and conceptual sense as well. Final decisions regarding the selection of factors and the items that loaded on them were made based on a combination of statistical fit and consistency with theory from the literature.

Multilevel confirmatory factor analyses. First, multilevel confirmatory factor analysis (MCFA; Muthén, 1994) procedures were conducted using the September/October 2007 sample. Results of the MEFA were used to construct the MCFA model as the second step to addressing *Research Question 1*. MCFA was conducted using Mplus Version 7.3 (Muthén & Muthén, 1998-2012). The first factor pattern coefficient (i.e., loading) was set to 1.0 to scale the educator and school-level factors. Models were based on the pooled within-group and between-group covariance matrices. Robust maximum likelihood estimation was used to determine the parameters of the models. Overall goodness of fit for each model was examined using the criteria for the MEFA described above. These analyses were designed to determine the extent to which factors derived at the educator-level were similar to or different from school-level factors.

Research has demonstrated that the factor structure at Level 2 often differs from Level 1 (Dedrick & Greenbaum, 2011).

Because factor labels and items that loaded on each factor were the same across the educator- and school-levels, the Satorra-Bentler chi-square difference test was used to determine whether the null hypothesis of equal factor loadings could be rejected (*Research Question 2*). Alpha was set at .05. Additionally, the BIC and saBIC for both the unconstrained model (the model derived from *Research Question 1*) and the equal loadings model were compared to determine which model indicated better fit (smaller numbers indicate better fit). The BIC and saBIC were chosen for comparison because they account for sample size and tend to be more effective in choosing the model that more accurately fits the parameters of the sample (Raffalovich, Glenn, Armstrong, & Hui-Shien, 2008).

The model derived from the MCFA and Satorra-Bentler chi-square difference test using the September/October 2007 sample was then analyzed using the data from the September/October 2008 sample and the January/February 2010 sample. In other words, the resulting model from the first MCFA was examined for goodness of fit at the remaining two time points. MCFA procedures and decision-making criteria were the same as for the September/October 2007 sample. Fit indices, factor structure, item loadings, and correlations among factors were compared across all three waves to determine the similarity of the factor structure across time (*Research Question 3*).

The model that resulted from the aforementioned analyses was then applied for each year to address *Research Question 5*. Specifically, the total score from the *Tier I and II Critical Components Checklists* was entered into the model at Level 2 to examine the relationship between the factor scores from the *Perceptions of Practices Survey* and another measure of MTSS implementation fidelity. Alpha was set at .05 for each wave of administration.

Reliability. Reliability analyses were conducted to address *Research Question 4* for all three waves. When analyzing reliability for multilevel nested data, it is necessary to account for the multilevel variability. To get more accurate reliability estimates when conducting an MCFA, composite reliabilities are calculated from factor loadings (Geldhof, Preacher, & Zyphur, 2014). This method is considered more accurate than Cronbach's alpha that is commonly used for single-level analyses. Thus, composite reliabilities were calculated using Mplus Version 7.3 (Muthén & Muthén, 1998-2012).

Endnotes

[2]Although Castillo et al., 2015 references the *Beliefs on RtI Survey*, the same EEVP procedures were used for the *Perceptions of Practices Survey*

CHAPTER IV: Results

Results related to the research questions are presented in this chapter. First the descriptive statistics, assumptions, and response analyses (i.e., analyses of the *Don't Know* responses) will be discussed followed by the results for each research question. The initial multilevel exploratory factor analysis (MEFA) that examined both academic and behavior items from the *Perceptions of Practices Survey* did not converge. Because all schools in the study focused on academics when implementing MTSS, it was decided that the academic items would be the focus of the current study. Thus, an MEFA with only academic items was conducted, which resulted in convergence. The results presented in this chapter will be related to academic items only.

Preliminary Analyses

Descriptive data. Item means for the September/October 2007 sample ranged from 3.73 (*SD* = 1.06 & 1.15) for items 14a (teacher receives staff support to implement intervention plan) and 16a (data were graphed routinely), respectively, to 4.45 (*SD* = 0.74) for item 4a (data were used to identify at-risk students in need of intervention). Item means for the September/October 2008 sample ranged from 3.81 (*SD* = 1.05 & 1.09) for items 14a and 16a, respectively, to 4.46 (*SD* = 0.72) for item 4a. Item means for the January/February 2010 sample ranged from 3.95 (*SD* = 1.04) for item 14a to 4.52 (*SD* = 0.67) for item 4a. See Tables 3-5 for all descriptive statistics for 2007, 2008, and 2010 samples.

Intraclass correlations (ICCs) for each of the items also were examined. ICCs close to zero indicate that nearly all variation is at the educator level, whereas ICCs close to 1.00 indicate that nearly all variation is at the school level. ICCs at or above .05 indicate that multi-level factor analysis is an appropriate method (Dedrick & Greenbaum, 2008; Dyer, Hanges, & Hall, 2005). ICCs for September/October 2007 ranged from .07 (item 3a, data used to make decisions about changes to core curriculum to increase student performance) to .14 (item 16a). ICCs for September/October 2008 ranged from .04 (item 10c academics, quantifiable data used to identify peer performance) to .11 (item 14a). ICCs for January/February 2010 ranged from .07 (item 2a, data used to identify students receiving core instruction that achieved benchmarks) to .15 (item 14a). See Tables 3-5 for all ICCs for 2007, 2008, and 2010 samples. Overall, the ICCs indicated sufficient variability between schools to conduct multilevel analyses.

Assumptions. The assumption of normality was evaluated using skewness and kurtosis values. Skewness and kurtosis values indicated that the data were approximately normally distributed for the majority of items. Skewness values for all three years of data ranged from -1.46 (*SE* = 0.05) for item 4a of the September/October 2007 sample to -0.54 (*SE* = 0.06) for item 14a of the September/October 2007 sample. Kurtosis values for all three years of data ranged from -0.40 (*SE* = 0.12) for item 16a of the September/October 2007 sample to 2.56 (*SE* = 0.11) for item 4a of the September/October 2007 sample; however, the majority of kurtosis values were below 2.00.

For the September/October 2007 sample, the percentage of missing data likely is an overestimate whereas the percentage of *Don't Know* responses likely is an underestimate due to a lack of differentiation between missing data and *Don't Know* responses during data entry. During data entry, some of the *Don't Know* responses were entered as missing data, which skewed the percentage of missing data. This error in data entry was corrected for the 2008 sample. Thus, the percentages for 2008 and 2010 are a more accurate representation of missing data and *Don't Know* responses. Overall, there was a small fraction of missing information and there was no pattern in the missing data indicating that the data were missing at random. Therefore, the assumption of randomly missing data did not appear to be violated.

***Don't Know* responses.** Significant correlations were found between items with *Don't Know* responses indicating that respondents that selected *Don't Know* on one item also selected *Don't Know* on other items. As mentioned previously, estimates for the September/October 2007 sample are skewed because of data entry errors, so *Don't Know* responses were not further analyzed. In September/October 2008, item 4a (data were used to identify at-risk students in need of intervention) had the least number of *Don't Know* responses (n = 100, 5.00%) while item 11a (Problem-Solving Team routinely developed hypotheses regarding causes of student not meeting benchmarks) had the most *Don't Know* responses (n = 431, 21.50%). For the January/February 2010 sample, item 5a (at-risk students routinely received additional interventions) had the least number of *Don't Know* responses (n = 69, 3.30%) and item 11a had the most *Don't Know* responses (n = 302, 14.50%). Overall, for the majority of items, the number of *Don't Know* responses did not exceed 22% per item for a group of respondents. Furthermore, when responses were examined by educator job position, there was no relationship found between position and number of *Don't Know* responses. Therefore, the *Don't Know* responses did not appear to be any substantial impact on the data.

Table 3

Descriptive Data for 2007 Sample (Level-1 n = 2,109, Level-2 n = 62)

Item	Mean	SD	Skewness	Kurtosis	ICC	% Missing	% Don't Know Responses
2a	4.38	0.75	-1.33	2.47	0.07	7.90	1.00
3a	4.16	0.87	-1.01	0.91	0.07	9.60	0.80
4a	4.45	0.74	-1.46	2.56	0.10	4.30	0.50
5a	4.31	0.79	-0.99	0.61	0.11	4.20	0.30
6a	4.36	0.81	-1.27	1.41	0.11	8.10	0.50
7a	4.30	0.83	-1.19	1.36	0.09	11.80	0.80
8a	3.94	0.99	-0.80	0.17	0.09	21.20	1.40
9a	4.14	0.89	-0.95	0.61	0.11	12.00	0.90
10a-academics	4.35	0.79	-1.24	1.75	0.08	9.10	0.90
10b-academics	4.35	0.77	-1.18	1.55	0.07	9.10	0.70
10c-academics	4.15	0.89	-1.00	0.77	0.09	16.30	0.90
11a	3.82	1.03	-0.65	-0.21	0.08	23.50	1.30
12a	4.08	0.94	-0.95	0.50	0.08	13.00	1.60
13a	4.13	0.92	-0.98	0.66	0.10	10.90	1.10
14a	3.73	1.06	-0.54	-0.39	0.09	12.40	1.00
15a	3.98	0.97	-0.76	0.04	0.09	13.50	1.20
16a	3.73	1.15	-0.65	-0.40	0.14	21.70	1.70
17a-academics	4.21	0.90	-1.13	1.15	0.12	10.40	0.80
17b-academics	4.12	0.94	-0.97	0.55	0.10	14.20	1.30
17c-academics	4.01	0.99	-0.91	0.40	0.11	18.70	1.70
18a	3.85	1.08	-0.78	-0.07	0.10	20.70	1.60

Table 4

Descriptive Data for 2008 Sample (Level-1 n = 1,940, Level-2 n = 61)

Item	Mean	SD	Skewness	Kurtosis	ICC	% Missing	% Don't Know Responses
2a	4.43	0.69	-1.14	1.76	0.08	0.70	8.40
3a	4.18	0.84	-1.00	1.15	0.08	0.70	9.20
4a	4.46	0.72	-1.35	2.15	0.07	0.20	5.00
5a	4.29	0.81	-1.12	1.18	0.06	0.40	6.10
6a	4.39	0.76	-1.18	1.36	0.10	1.50	8.00
7a	4.34	0.79	-1.11	1.03	0.10	0.50	10.70
8a	3.98	0.96	-0.91	0.62	0.09	0.60	19.10
9a	4.24	0.80	-1.01	1.22	0.06	1.40	12.10
10a-academics	4.40	0.72	-1.14	1.54	0.05	1.30	9.90
10b-academics	4.39	0.72	-1.12	1.40	0.05	1.50	10.30
10c-academics	4.27	0.80	-1.01	0.98	0.04	2.40	14.70
11a	3.88	0.99	-0.79	0.24	0.07	1.40	21.50
12a	4.08	0.94	-1.05	1.01	0.09	1.20	14.20
13a	4.18	0.86	-1.06	1.24	0.07	1.50	12.30
14a	3.81	1.05	-0.68	-0.11	0.11	1.10	13.20
15a	4.03	0.92	-0.87	0.57	0.08	1.10	14.70
16a	3.81	1.09	-0.78	0.02	0.04	1.10	20.40
17a-academics	4.21	0.85	-1.06	1.19	0.08	1.00	11.90
17b-academics	4.12	0.89	-0.94	0.75	0.08	1.10	14.60
17c-academics	4.03	0.93	-0.88	0.55	0.08	1.70	18.10
18a	3.89	1.02	-0.83	0.31	0.09	1.30	19.40

Table 5

Descriptive Data for 2010 Sample (Level-1 n = 2,058, Level-2 n = 60)

Item	Mean	SD	Skewness	Kurtosis	ICC	% Missing	% Don't Know Responses
2a	4.48	0.69	-1.37	2.27	0.07	0.70	4.70
3a	4.30	0.78	-1.08	1.28	0.10	0.70	5.50
4a	4.52	0.67	-1.39	2.30	0.09	0.40	3.40
5a	4.40	0.78	-1.38	2.30	0.10	0.40	3.30
6a	4.47	0.74	-1.39	1.85	0.08	0.90	4.20
7a	4.39	0.76	-1.15	1.10	0.09	0.60	6.90
8a	4.14	0.95	-1.08	0.85	0.09	0.90	13.80
9a	4.33	0.79	-1.04	0.75	0.06	1.80	8.20
10a-academics	4.45	0.71	-1.28	1.75	0.07	1.70	6.30
10b-academics	4.46	0.70	-1.26	1.70	0.09	2.00	6.20
10c-academics	4.36	0.78	-1.23	1.60	0.09	2.90	9.30
11a	4.09	0.92	-0.90	0.35	0.10	1.90	14.50
12a	4.25	0.87	-1.15	1.07	0.11	1.80	8.20
13a	4.27	0.84	-1.09	1.03	0.12	1.90	7.40
14a	3.95	1.04	-0.80	-0.01	0.15	1.60	8.60
15a	4.26	0.85	-0.99	0.49	0.13	1.70	8.00
16a	4.14	0.96	-1.04	0.61	0.15	1.70	10.60
17a-academics	4.38	0.76	-1.15	1.09	0.10	1.50	6.90
17b-academics	4.32	0.79	-1.07	0.84	0.10	1.70	8.10
17c-academics	4.27	0.84	-1.07	0.77	0.10	1.70	10.90
18a	4.07	0.99	-0.97	0.42	0.14	1.60	13.40

Research Question 1

Is the factor structure (i.e., number of factors and factor loadings) underlying the *Perceptions of Practices Survey* at the educator level similar to or different from the structure at the school level?

Multilevel exploratory factor analysis. Multilevel exploratory factor analyses (MEFA) were conducted using the September/October 2007 sample data. MEFAs were executed on all combinations of models with one to five factors on each level. All 25 solutions were considered and evaluated. Overall, the ranges for the fit indices were as follows: AIC from 75,643.88 to 79,451.40, BIC from 76,720.52 to 80,316.86, saBIC from 76,151.82 to 79,748.16, CFI from .82 to .98, SRMR-within from .02 to .06, SRMR-between from .03 to .09. All fit statistics for the models are presented in Table 6.

Comparison of fit indices for the models from the MEFA led to the identification of the models with five and four within factors having the best fit. These models were determined to have the best fit because the AIC, BIC, and saBIC were lower for these models than the others, CFI values were greater than .95, and SRMR and RMSEA values were less than or equal to .08 (Hu & Bentler, 1999). After these models were identified, factor loadings for each of the models were examined using a criterion of .30 to determine whether an item loaded on a given factor. Although the models with five within factors had the best fit indices, the factors did not make theoretical sense. Therefore, the models with four within factors were examined for feasibility and interpretability. For the September/October 2007 sample of educators (n = 2,109), four factors at the within level made the most conceptual sense. The four factors resulting from the MEFA were labeled: (a) *Tiers I & II Problem Solving* (items 2a, 3a, 4a, 5a, 6a, 7a, 8a), (b) *Tier III Problem Identification* (items 9a, 10a academics, 10b academics, 10c academics), (c) *Tier III Problem Analysis & Intervention Procedures* (items 11a, 12a, 13a, 14a), and (d) *Tier III Evaluation of Response to Intervention* (items 15a, 16a, 17a academics, 17b academics, 17c academics, 18a). The MEFA did not indicate that the items loaded similarly to these factors at the between level. A review of other possible factor solutions at the between level and unrestricted models with zero factors at the within level revealed that four factors at the between level had the best fit. However, the item loadings for the four factors at the between level did not make strong conceptual or theoretical sense. Because the items that loaded on the four factors at the within level made the most theoretical sense and were a good fit to the data, it was decided that the four factor solution found at the within level would be tested at the between level to determine if the factors fit the data at both levels.

To determine whether the four factors at the within level derived from the MEFA were also feasible at the between level, a multilevel confirmatory factor analysis (MCFA) was conducted using the September/October 2007 sample data. For the first MCFA model (Model 1), the four within factors from the MEFA were replicated and specified at both the between and within levels. The first model was a four-factor model with loadings freely estimated across the educator and school levels. This relaxed model indicated reasonable fit (CFI = .93, RMSEA = .04, SRMR-within = .04, SRMR-between = .07). A second four-factor model (Model 2) was examined to test the equality of factor loadings across levels. Loadings across Level 1 and Level 2 were constrained to be equal by imposing cross-level invariance. The equal loadings model also indicated reasonable fit (CFI = .93, RMSEA = .04, SRMR-within = .04, SRMR-between = .07). See Table 7 for fit indices for each of the models and see Tables 8-9 for the unstandardized factor loadings for each model.

Table 6

Two-Level Exploratory Factor Analyses Fit Statistics for 2007 Sample (Level-1 n = 2,109, Level-2 n = 62)

Within Level Factors	Between Level Factors	AIC	BIC	saBIC	χ^2 (df)	CFI	RMSEA	SRMR-(within)	SRMR-(between)
5	5	75,648.38	77,078.83	76,275.02	711.25 (230)	.98	.03	.02	.03
5	4	75,643.88	76,978.22	76,228.42	740.76 (247)	.98	.03	.02	.05
5	3	75,653.39	76,885.96	76,193.35	786.27 (265)	.98	.03	.02	.06
5	2	75,673.34	76,798.48	76,166.23	844.21 (284)	.98	.03	.02	.08
5	1	75,708.46	76,720.52	76,151.82	919.33 (304)	.98	.03	.02	.09
4	5	76,023.78	77,358.12	76,608.32	1,120.66 (247)	.96	.04	.03	.03
4	4	76,017.88	77,256.10	76,560.31	1,148.75 (264)	.96	.04	.03	.05
4	3	76,032.35	77,168.80	76,530.20	1,199.23 (282)	.96	.04	.03	.05
4	2	76,061.04	77,090.07	76,511.83	1,265.92 (301)	.96	.04	.03	.07
4	1	76,095.86	77,011.80	76,497.11	1,340.73 (321)	.96	.04	.03	.08
3	5	76,719.65	77,952.21	77,259.61	1,852.53 (265)	.94	.05	.03	.03
3	4	76,719.38	77,855.83	77,217.23	1,886.26 (282)	.93	.05	.03	.05
3	3	76,729.85	77,764.53	77,183.12	1,932.73 (300)	.93	.05	.03	.05
3	2	76,759.43	77,686.68	77,165.64	2,000.31 (319)	.93	.05	.03	.06
3	1	76,801.58	77,615.75	77,158.25	2,082.46 (339)	.93	.05	.03	.08
2	5	77,606.93	78,732.07	78,099.82	2,777.81 (284)	.90	.06	.05	.03
2	4	77,621.27	78,650.30	78,072.06	2,826.15 (301)	.90	.06	.05	.04
2	3	77,624.65	78,551.90	78,030.85	2,865.52 (319)	.90	.06	.04	.04
2	2	77,645.99	78,465.82	78,005.14	2,924.87 (338)	.89	.06	.04	.06
2	1	77,708.13	78,414.88	78,017.74	3,027.01 (358)	.89	.06	.05	.07
1	5	79,304.80	80,316.86	79,748.16	4,515.68 (304)	.83	.08	.06	.03
1	4	79,313.50	80,229.45	79,714.75	4,558.38 (321)	.83	.08	.06	.03
1	3	79,331.46	80,145.63	79,688.13	4,612.34 (339)	.82	.08	.06	.05
1	2	79,380.74	80,087.48	79,690.35	4,699.62 (358)	.82	.08	.06	.06
1	1	79,451.40	80,045.07	79,711.47	4,810.28 (378)	.82	.08	.06	.07

Note. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; saBIC = Sample-size Adjusted BIC; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual.

Table 7

Multilevel Confirmatory Factor Analysis Fit Indices

Model	AIC	BIC	saBIC	CFI	RMSEA	SRMR-within	SRMR-between
1: Four factors at Levels 1 & 2: Loadings freely estimated (Relaxed model)	76,578.20	77,239.71	76,867.99	.93	.04	.04	.07
2: Four factors at Levels 1 & 2: Cross-level invariance imposed (Constrained model)	76,584.26	77,149.66	76,831.95	.93	.04	.04	.07

Note. Level-1 n = 2,109, Level-2 n = 62.

Research Question 2

Are the factor loadings underlying the *Perceptions of Practices Survey* significantly different between and within schools?

The Satorra-Bentler chi-square difference test was used to determine whether the null hypothesis of equal factor loadings could be rejected. The $\Delta\chi^2$ was 33.78 (Δdf = 17, p < .01) indicating that loadings significantly differed across levels. The statistically significant difference in chi-square indicated that the baseline (loadings freely estimated) model fit the data better than the constrained comparison model (loadings fixed to be equal). However, the overall BIC and sample adjusted BIC were smaller for the constrained model (BIC = 77,149.66 & saBIC = 76,831.95) indicating a better fit for the equal loadings model. Because the constrained model indicated better fit according to the BIC and saBIC, the decision was made that the factor loadings were not significantly different across levels.

Table 8

Unstandardized Parameter Estimates for Model 1: Four Factors at Levels 1 & 2 with Loadings Freely Estimated (2007 Sample; Level-1 n = 2,109, Level-2 n = 62)

	Level 1: Educator					Level 2: School				
Item	TI & II PS	TIII PI	TIII PA & IX	TIII Eval	Residual Variance	TI & II PS	TIII PI	TIII PA & IX	TIII Eval	Residual Variance
2a	1.00[a] (—)				0.34 (0.03)	1.00[a] (—)				0.00 (0.00)
3a	1.34 (0.07)				0.38 (0.03)	1.16 (0.17)				0.01 (0.00)
4a	1.14 (0.06)				0.26 (0.03)	1.17 (0.18)				0.02 (0.01)
5a	1.23 (0.08)				0.28 (0.02)	1.34 (0.26)				0.01 (0.01)
6a	1.34 (0.09)				0.25 (0.02)	1.44 (0.33)				0.01 (0.00)
7a	1.49 (0.10)				0.22 (0.02)	1.41 (0.23)				0.00 (0.00)
8a	1.17 (0.09)				0.64 (0.04)	1.51 (0.32)				0.03 (0.01)
9a		1.00[a] (—)			0.47 (0.03)		1.00[a] (—)			0.03 (0.01)
10a academics		1.21 (0.07)			0.21 (0.02)		1.10 (0.27)			0.00 (0.00)
10b academics		1.26 (0.09)			0.16 (0.02)		0.93 (0.18)			0.00(0.00)
10c academics		1.32 (0.08)			0.30 (0.02)		1.21 (0.27)			0.01 (0.00)
11a			1.00[a] (—)		0.45 (0.03)			1.00[a] (—)		0.02 (0.01)
12a			1.08 (0.04)		0.24 (0.02)			1.01 (0.12)		0.00 (0.00)
13a			1.01 (0.04)		0.24 (0.02)			1.28 (0.23)		0.00 (0.00)
14a			0.86 (0.04)		0.64 (0.04)			1.28 (0.42)		0.03 (0.01)
15a				1.00[a] (—)	0.38 (0.02)				1.00[a] (—)	0.00 (0.00)
16a				1.02 (0.04)	0.61 (0.04)				1.05 (0.20)	0.10 (0.02)
17a academics				1.07 (0.04)	0.18 (0.02)				1.02 (0.09)	0.00 (0.00)
17b academics				1.19 (0.04)	0.12 (0.01)				0.96 (0.09)	0.00 (0.00)
17c academics				1.20 (0.05)	0.19 (0.02)				1.05 (0.11)	0.00 (0.00)
18a				1.08 (0.04)	0.47 (0.04)				1.11 (0.11)	0.00 (0.01)

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. [a]Factor loading fixed to 1.00.
Standard errors are in parentheses.
All items loaded at the .05 level.

Table 9

Unstandardized Parameter Estimates for Model 2: Four Factors at Levels 1 & 2 with Loadings Constrained to be Equal (2007 Sample; Level-1 n = 2,109, Level-2 n = 62)

	Levels 1 & 2: Educator & School					
Item	TI & II PS	TIII PI	TIII PA & IX	TIII Eval	Level 1 Residual Variance	Level 2 Residual Variance
2a	1.00[a] (—)				0.34 (0.03)	0.00 (0.00)
3a	1.33 (0.07)				0.38 (0.03)	0.01 (0.00)
4a	1.14 (0.06)				0.25 (0.03)	0.02 (0.01)
5a	1.24 (0.08)				0.28 (0.02)	0.01 (0.00)
6a	1.35 (0.09)				0.25 (0.02)	0.01 (0.00)
7a	1.48 (0.10)				0.22 (0.02)	0.00 (0.00)
8a	1.20 (0.09)				0.63 (0.05)	0.03 (0.01)
9a		1.00[a] (—)			0.47 (0.03)	0.03 (0.01)
10a academics		1.20 (0.08)			0.21 (0.02)	0.00 (0.00)
10b academics		1.23 (0.09)			0.16 (0.02)	0.00 (0.00)
10c academics		1.31 (0.09)			0.30 (0.02)	0.01 (0.00)
11a			1.00[a] (—)		0.45 (0.03)	0.02 (0.01)
12a			1.07 (0.04)		0.24 (0.02)	0.01 (0.00)
13a			1.02 (0.04)		0.24 (0.02)	0.00 (0.00)
14a			0.88 (0.04)		0.64 (0.04)	0.04 (0.02)
15a				1.00[a] (—)	0.37 (0.02)	0.01 (0.00)
16a				1.02 (0.04)	0.61 (0.04)	0.11 (0.02)
17a academics				1.06 (0.03)	0.18 (0.02)	0.00 (0.00)
17b academics				1.16 (0.03)	0.13 (0.01)	0.00 (0.00)
17c academics				1.19 (0.04)	0.19 (0.02)	0.00 (0.00)
18a				1.09 (0.04)	0.47 (0.04)	0.01 (0.01)

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All items loaded at the .05 level.
[a]Factor loading fixed at 1.00.

Research Question 3

To what extent are the factor structure, factor loadings, and correlations of factors different across multiple time points for the *Perceptions of Practices Survey*?

The constrained model demonstrated good fit for the 2007 (educator-level *n* = 2,109; school-level *n* = 62), 2008 (educator-level *n* = 1,940; school-level *n* = 61), and 2010 (educator-level *n* = 2,058; school-level *n* = 60) samples (CFI ranged from .92 to.93, RMSEA ranged from .04 to .05, SRMR-within ranged from .04 to .05, SRMR-between ranged from .06 to .11). See Table 10 for fit indices across time points. Unstandardized item loadings for 2007 ranged from 0.88 (SE = 0.04) to 1.48 (SE = 0.10). For 2008, item loadings ranged from 0.95 (SE = 0.02) to 1.51 (SE = 0.08). Item loadings for 2010 ranged from 0.85 (SE = 0.04) to 1.33 (SE = 0.06). All items significantly loaded on the respective factors. See Tables 9, 11, and 12 for unstandardized parameter estimates for 2007, 2008, and 2010, respectively. The correlations among factors were large across years. Correlations ranged from r = .65 (*Tier III Evaluation* with *Tier III Problem Identification* at the educator level) to r = 1.00 (*Tier III Problem Identification* with *Tier III Problem Analysis and Intervention Procedures* at the school level). The majority of correlations fell between r = .65 and r = .98. See Tables 13, 14, and 15 for factor correlations for 2007, 2008, and 2010 respectively.

Due to the high correlations among factors found at the between-level, it was decided to test alternative models with one between-level factor and three between-level factors. The model with three between-level factors combined the two highest correlated factors: *Tier III Problem Identification* and *Tier III Problem Analysis & Intervention Procedures*. Because the model with one between-level factor (CFI = .93, RMSEA = .04, SRMR-within = .04, SRMR-between = .08) and the model with three between-level factors (CFI = .93, RMSEA = .04, SRMR-within = .04, SRMR-between = .08) adequately fit the data, the two alternative models were compared with the four between-level factor model using the Satorra-Bentler chi-square difference test. The $\Delta\chi^2$ indicated that both the one-factor ($\Delta\chi^2$ = 37.86, Δdf = 6, p < .01) and four-factor models ($\Delta\chi^2$ = 15.23, Δdf = 3, p < .01) were significantly different from the three-factor model. However, the overall BIC (four-between = 77,149.66, three-between = 77,152.68, one-between = 77,149.49) and saBIC (four-between = 76,831.95, three-between = 76,844.50, one-between = 76,850.84) indicated that the four between-level factor model fit the data better than the alternative models.

Table 10

Model Fit Indices Across Time Points

Sample	AIC	BIC	saBIC	CFI	RMSEA	SRMR-within	SRMR-between
September/ October 2007	76,584.26	77,149.66	76,831.95	.93	.04	.04	.07
September/ October 2008	66,485.23	67,042.28	66,724.58	.92	.05	.05	.11
January/ February 2010	64,293.38	64,856.33	64,538.62	.93	.05	.04	.06

Note. 2007: Level-1 n = 2,109, Level-2 n = 62. 2008: Level-1 n = 1,940, Level-2 n = 61. 2010: Level-1 n = 2,058, Level-2 n = 60.

Research Question 4

What is the reliability of the scores from the *Perceptions of Practices Survey* at the educator and school levels across multiple time points?

Composite reliability estimates were used to examine reliability at the educator and school levels. Reliability estimates at the educator level for the factors of *Tiers I & II Problem Solving and Intervention*, *Tier III Problem Identification*, *Tier III Problem Analysis & Intervention Procedures*, and *Tier III Evaluation of Response to Intervention* for the September/October 2007 sample were .86, .84, .84, and .92, respectively. At the school-level, estimates for 2007 were .97, .94, .95, and .97, respectively. For the September/October 2008 sample, internal consistency estimates were .86, .89, .86, and .91, respectively. For the 2008 sample, composite reliability estimates at the school-level were .98, .98, .94, and .96, respectively. Finally, for the January/February 2010 sample, internal consistency estimates were .89, .90, .85, and .92, respectively. At the school-level, estimates for 2010 were .98, .98, .98, and .99. Thus, all factors across three years exceeded the .70 estimate typically considered acceptable for internal consistency reliability (Nunnally, 1978).

Research Question 5

To what extent are the factor scores derived from the *Perceptions of Practices Survey* related to the *Tiers I and II Critical Components Checklist*?

Level 2 (school) covariates were added to the constrained four-factor model to examine the relationship between school-level perceptions of practices and school-level implementation fidelity. Relationships between the total implementation score derived from the *Tiers I and II Critical Components Checklist* and perceptions of practices scores were investigated across multiple time points. From 2007-2010, the average *Total Problem Solving* score increased from 0.55 to 0.95. A higher score indicates higher levels of implementation. See Table 16 for descriptive statistics for this score across years. For the 2007 sample, parameter estimates between factors on the *Perceptions of Practices Survey* and the Total Problem Solving score on the *Tiers I and II Critical Components Checklist* ranged from -0.04 (p = 0.77) to 0.00 (p = 0.98). For 2008, estimates ranged from 0.15 (p = 0.31) to 0.22 (p = 0.11). Finally, for 2010, estimates ranged from 0.35 ($p < 0.05$) to 0.39 ($p < 0.05$). Only estimates for the 2010 sample significantly contributed to the model. See Table 17 for standardized parameter estimates between the factors on *Perceptions of Practices Survey* and the Total Problem Solving score from the *Tiers I & II Critical Components Checklist* across the three years.

Table 11

Unstandardized Parameter Estimates for 2008 (Level-1 n = 1,940, Level-2 n = 61)

	Levels 1 & 2: Educator & School					
Item	TI & II PS	TIII PI	TIII PA & IX	TIII Eval	Level 1 Residual Variance	Level 2 Residual Variance
2a	1.00[a] (—)				0.26 (0.02)	0.00 (0.00)
3a	1.26 (0.05)				0.30 (0.02)	0.01 (0.00)
4a	1.21 (0.05)				0.17 (0.02)	0.00 (0.00)
5a	1.34 (0.06)				0.22 (0.02)	0.00 (0.00)
6a	1.35 (0.07)				0.19 (0.01)	0.01 (0.00)
7a	1.51 (0.08)				0.16 (0.02)	0.00 (0.00)
8a	1.23 (0.07)				0.48 (0.04)	0.02 (0.01)
9a		1.00[a] (—)			0.34 (0.03)	0.00 ((0.00)
10a academics		1.24 (0.05)			0.08 (0.01)	0.00 (0.00)
10b academics		1.24 (0.05)			0.07 (0.01)	0.00 (0.00)
10c academics		1.26 (0.04)			0.17 (0.02)	0.00 (0.00)
11a			1.00[a] (—)		0.32 (0.02)	0.02 (0.01)
12a			0.99 (0.03)		0.25 (0.02)	0.00 (0.00)
13a			0.98 (0.03)		0.21 (0.02)	0.00 (0.00)
14a			0.97 (0.03)		0.51 (0.04)	0.03 (0.01)
15a				1.00[a] (—)	0.21 (0.03)	0.00 (0.01)
16a				1.03 (0.03)	0.33 (0.03)	0.02 (0.01)
17a academics				0.95 (0.02)	0.13 (0.01)	0.00 (0.00)
17b academics				0.98 (0.03)	0.14 (0.01)	0.00 (0.00)
17c academics				1.05 (0.03)	0.16 (0.02)	0.00 (0.00)
18a				1.07 (0.04)	0.41 (0.04)	0.00 (0.01)

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All items loaded at the .05 level. [a]Factor loading fixed to 1.00.

Table 12

Unstandardized Parameter Estimates for 2010 (Level-1 n = 2,058, Level-2 n = 60)

Item	Levels 1 & 2: Educator & School				Level 1 Residual Variance	Level 2 Residual Variance
	TI & II PS	TIII PI	TIII PA & IX	TIII Eval		
2a	1.00[a] (—)				0.27 (.02)	0.00 (0.00)
3a	1.16 (0.04)				0.38 (0.03)	0.01 (0.01)
4a	1.09 (0.04)				0.22 (0.02)	0.00 (0.00)
5a	1.27 (0.05)				0.31 (0.02)	0.00 (0.00)
6a	1.22 (0.06)				0.21 (0.02)	0.01 (0.00)
7a	1.33 (0.06)				0.17 (0.01)	0.00 (0.00)
8a	1.30 (0.07)				0.60 (0.04)	0.02 (0.01)
9a		1.00[a] (—)			0.35 (0.02)	0.01 (0.00)
10a academics		1.22 (0.05)			0.11 (0.01)	0.00 (0.00)
10b academics		1.28 (0.05)			0.08 (0.01)	0.00 (0.00)
10c academics		1.27 (0.05)			0.19 (0.02)	0.01 (0.00)
11a			1.00[a] (—)		0.36 (0.03)	0.02 (0.01)
12a			1.03 (0.03)		0.22 (0.03)	0.01 (0.00)
13a			0.96 (0.03)		0.19 (0.02)	0.00 (0.00)
14a			0.85 (0.04)		0.60 (0.05)	0.05 (0.01)
15a				1.00[a] (—)	0.37 (0.04)	0.01 (0.01)
16a				1.07 (0.04)	0.63 (0.05)	0.07 (0.02)
17a academics				1.10 (0.04)	0.17 (0.02)	0.00 (0.00)
17b academics				1.24 (0.04)	0.10 (0.01)	0.00 (0.00)
17c academics				1.24 (0.04)	0.16 (0.02)	0.00 (0.00)
18a				1.07 (0.03)	0.46 (0.04)	0.02 (0.01)

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All items loaded at the .05 level.
[a]Factor loading fixed to 1.00.

Table 13

Standardized Factor Correlations for 2007 (Level-1 n = 2,109, Level-2 n = 62)

		TI & II PS	TIII PI	TIII PA & IX	TIII Eval
Level 1: Educato	TI & II PS				
	TIII PI	.79 (0.02)			
	TIII PA & IX	.73 (0.02)	.75 (0.02)		
	TIII Eval	.73 (0.02)	.76 (0.02)	.79 (0.02)	
Level 2: School	TI & II PS				
	TIII PI	.98 (0.03)			
	TIII PA & IX	.94 (0.04)	1.00 (0.03)		
	TIII Eval	.96 (0.03)	.93 (0.04)	.95 (0.04)	

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All correlations significant at .05 level.

Table 14

Standardized Factor Correlations for 2008 (Level-1 n = 1,940, Level-2 n = 61)

		TI & II PS	TIII PI	TIII PA & IX	TIII Eval
Level 1: Educato	TI & II PS				
	TIII PI	.78 (0.03)			
	TIII PA & IX	.71 (0.02)	.71 (0.03)		
	TIII Eval	.68 (0.02)	.65 (0.03)	.75 (0.02)	
Level 2: School	TI & II PS				
	TIII PI	.94 (0.05)			
	TIII PA & IX	.92 (0.09)	.79 (0.14)		
	TIII Eval	.96 (0.04)	.92 (0.06)	.90 (0.08)	

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All correlations significant at .05 level.

Table 15

Standardized Factor Correlations for 2010 (Level-1 n = 2,058, Level-2 n = 60)

		TI & II PS	TIII PI	TIII PA & IX	TIII Eval
Level 1: Educato	TI & II PS				
	TIII PI	.80 (0.02)			
	TIII PA & IX	.78 (0.02)	.77 (0.02)		
	TIII Eval	.77 (0.02)	.77 (0.02)	.80 (0.02)	
Level 2: School	TI & II PS				
	TIII PI	.98 (0.02)			
	TIII PA & IX	.92 (0.03)	.94 (0.03)		
	TIII Eval	.94 (0.03)	.95 (0.03)	.97 (0.03)	

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention. Standard errors are in parentheses. All correlations significant at .05 level.

Table 16

Descriptive Data for Total Problem Solving Scores Across Years

Year	*N*	*M*	*SD*	Minimum	Maximum	Skewness	Kurtosis
2007	2,130	0.55	0.47	0.00	1.50	0.62	-0.87
2008	2,001	0.64	0.50	0.00	1.73	0.44	-0.85
2010	2,088	0.95	0.61	0.00	1.94	0.06	-1.24

Note. Average implementation scores for educators nested within schools.

Table 17

Standardized Parameter Estimates Between Factors on Perceptions of Practices Survey with Tiers I & II Critical Components Checklist

Year	TI & II PS with TotPS	Tier III PI with TotPS	Tier III PA & IX with TotPS	Tier III Eval with TotPS
2007 (Level-1 *n* = 2,109, Level-2 *n* = 62)	0.00 (0.13)	-0.02 (0.14)	-0.02 (0.14)	-0.04 (0.13)
2008 (Level-1 *n* = 1,940, Level-2 *n* = 61)	0.21 (0.12)	0.15 (0.14)	0.22 (0.13)	0.21 (0.14)
2010 (Level-1 *n* = 2,058, Level-2 *n* = 60)	0.35* (0.14)	0.36* (0.15)	0.39* (0.13)	0.35* (0.13)

Note. TI & II PS = Tiers I & II Problem Solving, TIII PI = Tier III Problem Identification, TIII PA & IX = Tier III Problem Analysis & Intervention Procedures, TIII Eval = Tier III Evaluation of Response to Intervention, TotPS = Total Problem Solving. Standard errors are in parentheses.

**p* < .05.

CHAPTER V: Discussion

The purpose of this study was to examine the reliability and validity of the *Perceptions of Practices Survey* for use in measuring educators' self-reported MTSS implementation within schools. This chapter presents a discussion of the findings, limitations of the current study, and implications for research and practice.

Discussion of Findings

As the adoption of MTSS has been increasing, researchers have been developing tools to measure the fidelity of implementation and to evaluate the effectiveness of MTSS. One of these tools is the *Perceptions of Practices Survey*, a self-report survey that can be completed by school staff to measure their perceptions of MTSS practices within a school. A prior study examined the factor structure of the tool using single-level EFA procedures (Castillo et al., 2010); however, the current study extended the literature by using multilevel statistical analyses that accounted for the nested data structure (i.e., educators nested within schools). Results of the MEFA and MCFAs examining academic items only across multiple years supported a four-factor structure at both the educator and school levels. The four factors were labeled: (a) *Tiers I & II Problem Solving* (items 2a, 3a, 4a, 5a, 6a, 7a, 8a), (b) *Tier III Problem Identification* (items 9a, 10a academics, 10b academics, 10c academics), (c) *Tier III Problem Analysis & Intervention Procedures* (items 11a, 12a, 13a, 14a), and (d) *Tier III Evaluation of Response to Intervention* (items 15a, 16a, 17a academics, 17b academics, 17c academics, 18a).

The four-factor solution found in the current study is consistent with the literature on MTSS and problem-solving. An MTSS commonly is organized around three levels of student support and the problem-solving process often is used to make decisions about instruction and intervention (Batsche et al., 2007; Burns & VanDerHeyden, 2006; Fletcher & Vaughn, 2009; Stoiber, 2014). A common problem-solving framework includes four steps: (1) Problem Identification (i.e., well-defined problem that includes the discrepancy between what is occurring and what is expected); (2) Problem Analysis (i.e., hypotheses regarding why the problem is occurring that are either confirmed or rejected based on data); (3) Intervention Plan Development and Implementation (i.e., a plan is developed and implemented that includes evidence-based strategies that are matched to needs determined by validated hypotheses); and (4) Evaluation of the response to intervention (i.e., a decision regarding how much progress was made based on frequently collected data; Bergan & Kratochwill, 1990; Gutkin & Curtis, 2009).

The resulting factor structure from the current study aligns with the practices of applying the problem-solving framework across multiple tiers (Batsche et al., 2007; Brown-Chidsey & Steege, 2010; Burns & VanDerHeyden, 2006; Fletcher & Vaughn, 2009; Pluymert, 2014; Stoiber, 2014). The first seven items on the *Perceptions of Practices Survey* require educators to indicate how often problem-solving practices occur at Tiers I & II (see Table 18), resulting in the labeling of that factor as *Tiers I & II Problem-Solving*. The goal of problem-solving at Tier I is to determine how students are progressing within the core curriculum and what adjustments need to be made to improve student performance. School problem-solving teams gather universal screening data (e.g., Curriculum-Based Measures, state assessments) to determine the percentage of students not making progress in the core curriculum and then the team analyzes the data to determine the changes or interventions that are needed to improve student performance (Batsche et al., 2005; Pluymert, 2014; Stoiber, 2014; Tilly, 2008). The application of these problem-solving steps at Tier I are reflected in questions 2a and 3a on the survey. At the next tier, Tier II, the problem-solving team uses school-wide data to determine which students are at-risk and what interventions these students need based on their presenting problems. These procedures at Tier II are reflected by items 4a and 5a on the survey. Next, the team continues to collect data to monitor the at-risk students' progress and to determine the percentage of students who received the supplemental interventions and achieved grade-level benchmarks. These steps are measured by items 6a and 7a. Item 8a on the survey asks educators if a "standard protocol intervention (i.e., the same type of intervention used for similar problems" [Castillo, Hines, Batsche, & Curtis, 2011; p. 2]) was used for all at-risk students. The use of a standard protocol is a common procedure for Tier II interventions in a school (Hughes & Dexter, 2011), thus making it a reasonable addition to the factor *Tiers I & II Problem-Solving*.

The remaining items on the survey focus on the problem-solving steps applied to Tier III (see Table 18). Items 9a through 10c academics on the survey loaded onto the factor labeled *Tier III Problem Identification*. This label was given to these items because they outline specific steps the problem-solving team should follow to define a student's target behavior including defining the problem in terms of the desired behavior and using quantifiable data to determine the student's current performance level, desired performance level, and his/her performance in comparison to same-age peers (Batsche et al., 2005; Pluymert, 2014; Stoiber, 2014; Tilly, 2008).

Items 11a through 14a loaded on the factor labeled *Tier III Problem Analysis and Intervention Procedures*. These items are believed to have loaded on the same factor because the steps that are required for intervention procedures are dependent on the steps for problem analysis (Batsche et al., 2005; Pluymert, 2014; Stoiber, 2014; Tilly, 2008). For instance, during problem analysis, the team develops hypotheses to explain why the student is not meeting expectations and gathers data to confirm or reject the hypotheses (reflected by items 11a and 12a on the survey). Then, for intervention development, the team develops interventions based on the confirmed hypotheses (item 13a). As part of the intervention development, it is best practice for the team to also plan for intervention support for the teacher implementing the intervention (reflected in item 14a).

The final factor, *Tier III Evaluation of Response to Intervention*, encompasses items 15a through 18a. These items were thought to load on the same factor because they measure components of evaluation, the final step of the problem-solving process (Batsche et al., 2005; Pluymert, 2014; Stoiber, 2014; Tilly, 2008). Item 15a measures the degree to which the fidelity of intervention implementation occurs while items 16a through 17c academics measure the extent to which educators graph student data and use the data to determine the student's rate of progress as well as the change in the gap between his/her performance with expected performance and peer performance. The last item, 18a, asks how often the student's response to intervention data are used to determine whether or not s/he has a disability, which is consistent with requirements for and current application of response to intervention data within the school setting (IDEIA, 2004; Spectrum K12, 2011).

The finding that the factor structure was similar at the educator- and school-levels also is consistent with the literature. Scholars indicate that the steps of the problem-solving process are similar at both levels because it is a universal process employed to solve various problems in various contexts (e.g., Batsche et al., 2005; Bergan & Kratochwill, 1990, Brown-Chidsey & Steege, 2010; Pluymert, 2014; Stoiber, 2014). Although there are factors that may influence the implementation of the problem-solving process at different levels (e.g., knowledge of educators employing the process, data available, resources, etc.), there is no reason to suspect that the constructs would differ across levels. The steps of the problem-solving process would be implemented similarly regardless of whether one educator or groups of educators within a school engaged in problem-solving. For example, whether defining a problem (step one of the problem-solving process) at the school (e.g., school teams examining student performance) or educator (e.g., individual teacher engaging in problem-solving for his/her students) level, the problem is clearly defined in terms of the desired behavior and the current and expected levels of performance are identified.

Table 18

Factors of Perceptions of Practices Survey and Focus of Practice Statements

Perceptions of Practices Survey factors	Focus of practice statements
Tiers I & II Problem Solving (items 2a-8a)	Use of data to evaluate core effectiveness; data-based decision making for core curriculum; use of data to identify at-risk students; supplemental interventions for at-risk students; progress monitoring of at-risk students and data used to determine student achievement; standard intervention protocol used for at-risk students
Tier III Problem Identification (items 9a-10c academics)	Desired behavior used to define target behavior; quantifiable data used to identify target student's and peer's current and desired levels of performance
Tier III Problem Analysis & Intervention Procedures (items 11a-14a)	Hypotheses routinely developed; data used to validate hypotheses; intervention plans developed based on hypotheses; intervention implementation support
Tier III Evaluation of Response to Intervention (items 15a-18a)	Intervention fidelity data collected routinely; Data graphed routinely; progress monitoring data used to determine target student's rate of progress, decrease in gap of performance between current and desired levels and peers' performance

Of note is the fact that the model demonstrated good fit across the 2007, 2008, and 2010 samples of educators. Moreover, the factor loadings and correlations among factors were fairly similar across the three years, and the scores derived for the educator- and school-level factors provided reliable information. It was expected that the model would fit across years because the problem-solving process is a stable construct that is unlikely to change over time. Despite some variations to the steps included in the problem-solving process (e.g. specific procedures), the basic foundations of the process remain constant regardless of time (Bergan & Kratochwill, 1990; Gutkin & Curtis, 2009). Although the components of the problem-solving process (e.g., problem identification, problem analysis, intervention implementation, intervention evaluation) remain stable, it is expected that implementation levels of the various components would fluctuate based on professional development activities, leadership support, school climate, policies, procedures, and other system level factors (Fullan, 2010; Hall & Hord, 2015).

Finally, the constructs found in the current study were related to another measure of implementation. Implementation as measured by the *Tiers I and II Critical Components Checklist* was not significantly related to any of the factors on the *Perceptions of Practices Survey* for the 2007 and 2008 samples. However, the *Checklist* was significantly related to all four factors on the *Perceptions of Practices Survey* in 2010. This finding indicates that educators' perceptions of practices after three years of implementation were significantly related to trained and experienced MTSS coaches' ratings of problem-solving practices within a school, thereby providing evidence for the concurrent validity of the tool. One potential explanation for the lack of a significant relationship for the first two waves of administration involves the amount of variability in implementation scores from the *Checklist*. During the third year of the study, there was more variability in the *Checklist* scores, which likely increased the ability to detect relationships between the measures. Researchers argue that implementation of any innovation typically takes years before implementation with fidelity occurs (Batsche et al., 2005; Castillo & Curtis, 2014; Hall & Hord, 2015). Thus, it is plausible that relationships between the *Perceptions of Practices Survey* and other measures of implementation may not be detected until sufficient levels of implementation have been reached.

Another possible explanation for the absence of a significant relationship between the measures in 2007 and 2008 is that the average educator responses on the *Perceptions of Practices Survey* were high during those years (i.e., item response averages exceeded 4.0). These estimates were not commensurate implementation levels evidenced by the *Checklist* scores during 2007 and 2008. Thus, the educators' self-reported perceptions of practices in the first two years were potentially an upwardly biased estimate of practices (Desimone, 2009; Noell & Gansle, 2006). Perhaps educators believed problem-solving practices were occurring frequently in the school and were attempting to engage in the steps, but the process was not being implemented with enough fidelity to be evident in the permanent products that were evaluated to complete the *Checklists.*

Finally, when considering the relationships between the *Perceptions of Practices Survey* and the *Checklist*, it is important to consider the psychometric properties and limitations of the *Checklist*. Although the *Checklist* had good internal consistency (estimates using Cronbach's alpha exceeded .80) and evidence of high inter-rater agreement (91.16%), potential limitations inherent to permanent product reviews exist. Because the *Checklist* is completed based on the availability and quality of permanent products, the validity and accuracy of the tool may be affected by factors such as variability in access to and the status of documentation. For this reason, it is recommended that both tools be used in conjunction with other measures (e.g., observation protocols) to ensure accuracy.

Limitations and Future Research

A few limitations of the current study must be considered when interpreting the findings. One limitation involves the limited sample. Although the sample is representative of the population of educators from a large and diverse southeastern state, the number of schools was relatively small. For multilevel research, it is recommended to have at least 30-50 level two units (i.e., schools; Muthén & Muthén, n.d.) in order to achieve adequate power (Snijders, 2005) to better determine sources of model misfit. Although the current sample (Level-2 $n \approx 61$) met the sample size criterion for multilevel research, the inclusion of more schools in the sample would have increased the chance of finding a model that best fit the data.

Additionally, the state from which the samples were drawn had state requirements for the implementation of MTSS since 2007, which could have influenced the presence of MTSS practices. Because states vary in their policies regarding requirements for the implementation of MTSS (Hauerwas et al., 2013) and because MTSS models can vary in terms of processes and procedures (Fletcher & Vaughn, 2009), future research should include educators from various states to examine the generalizability of these findings to other settings with different expectations for and resources available to support MTSS implementation. Future research also should include a more geographically diverse sample to address how well the measure generalizes to other regions.

Another limitation of this study is that the sample only included elementary schools. Additional research should examine the evidence for the reliability and validity of the *Perceptions of Practices Survey* at the secondary level. Because of the differences between the two settings (e.g., time for interventions, availability of assessments, composition of problem-solving team), problem-solving processes may be perceived differently at the secondary level (Sansosti, Noltemeyer, & Goss, 2010). Although it is hypothesized that the four-factor structure would fit the data at the secondary level because the problem-solving process is a universal process that does not tend to greatly fluctuate (Bergan & Kratochwill, 1990; Gutkin & Curtis, 2009), future research is needed.

An additional limitation is that the current study only examined the academic items and not the behavior items on the *Perceptions of Practices Survey*. As mentioned earlier, the problem-solving process is a general practice that can be applied across content areas (e.g., Brown-Chidsey & Steege, 2010; Pluymert, 2014; Stoiber, 2014). Even though MTSS is a system that is intended to encompass academic and behavior problems, distinctive differences may exist between practices for academics and behavior (e.g., data sources, intervention practices; Stoiber, 2014). Therefore, future research is needed to examine the psychometric properties of the behavior items to confirm that the factor structure remains the same for these items.

Significantly high correlations among the factors at the between level indicate that the four factors may not be distinguishably different. Follow-up analyses in the current study revealed that the one and three between-level factor models adequately fit the data, and were significantly different from the four between-level factor model. However, based on the BIC and saBIC, the four between-level factor model better fit the data. The four-factor model also aligns with the problem-solving process found in the literature as well as training protocols for the problem-solving process. Moreover, when evaluating MTSS implementation, it is common to find that problem identification is being implemented more often than problem analysis and intervention procedures (Castillo, Hines, Batsche, & Curtis, 2011). Thus, for professional learning purposes, it would be beneficial to have separate factors to monitor to inform what steps of the problem-solving process require additional training and support. Nonetheless, future research should examine whether the four-factor model at the between level best explains the variance in school-level perceptions of practices.

The lack of clarity regarding the amount of *Don't Know* responses from the 2007 administration represents another limitation. Despite the uncertainty, *Don't Know* responses did not seem to have a substantial impact on the data. In general, the percentage of educators who responded with *Don't Know* was did not exceed 21% and on average were 13% for 2008 and 8% for 2010. When responses were examined by educator job position, there was no relationship found between position and number of *Don't Know* responses. Future research should include an analysis of *Don't Know* responses to determine their impact on survey results, especially when investigating the relationships between educator characteristics (e.g., job position, years of experience) and problem-solving practices. It is plausible that educators with different roles and responsibilities may have different understandings of the problem-solving practices occurring in their schools.

Future research also should seek to investigate the relationship between MTSS implementation as measured by the *Perceptions of Practices Survey* and important student outcomes. Researchers suggest that when MTSS is implemented with fidelity, schools and districts see improvements in reading and mathematics outcomes, academically related behaviors, general academic performance, and decreases in retention, referral and special education placement rates (Burns et al., 2005; Griffiths et al., 2007; Hughes & Dexter, 2011). Thus, relationships between scores from the survey and identified student outcomes would provide additional evidence for the validity of the tool by demonstrating associations found in studies using other implementation measures.

Implications for Practice

The current study extends the MTSS literature by providing evidence of reliability and validity for a self-report scale that measures educators' perceptions of MTSS practices within their schools. This is a noteworthy addition to the literature due to the limited availability of self-report tools that can be completed by multiple stakeholders within a school. Many existing tools require a select few educators on a leadership team to complete items (e.g., *TIC*; Sugai, Horner, Lewis-Palmer, Rossetto, 2012; *SET*; Sugai, Lewis-Palmer, Todd, & Horner, 2001; *BoQ*; Kincaid, Childs, & George, 2010; *EBS*; Sugai, Horner, & Todd, 2003; *Tier I and II Observation Checklist*; Castillo et al., 2010; *Tier I and II Critical Components Checklist, Tier III Critical Components Checklist;* Castillo et al., 2010). On the other hand, data from the *Perceptions of Practices Survey* can be efficiently gathered from multiple stakeholders to capture diverse perspectives regarding the problem-solving practices occurring within a school. The ability to capture perspectives of all school staff across grade-levels, roles and responsibilities, and content areas is important because problem-solving occurs in many settings and leadership team members (i.e., administration, instructional coaches, support personnel) are not always present. Because leadership team members are not always present in every problem-solving situation in the school, their perspectives may not reflect all aspects of the problem-solving practices occurring.

In addition to providing perspectives from stakeholders across a school, practitioners also may want to consider the potential utility of a self-report measure that can be used to monitor the fidelity of MTSS implementation within schools. Fidelity is a critical component of implementation to measure because if fidelity is not monitored, it is difficult to truly determine if improvements in student performance or a lack thereof are related to the implementation of MTSS or not (Carroll et al., 2007). Self-report measures are more efficient in terms of use of resources than other methods such as direct observations, interviews, and permanent products (Desimone, 2009). Because of concerns regarding the potential for upward bias in responses, researchers recommend that a combination of methods be used to measure fidelity (Desimone, 2009; Noell & Gansle, 2006). Thus, schools or districts can use information from the *Perceptions of Practices Survey* in conjunction with data gathered through other methods such as direct observations, interviews, or permanent product reviews to determine goals for professional development and evaluate growth in implementation fidelity (Schmoker, 2006). Using a combination of methods may be particularly important given the lack of relationship between the survey and the *Checklist* during the first two waves analyzed and the notion that the lack of relationships may have been influenced by higher than expected scores from the *Perceptions of Practices Survey*.

As part of a battery of tools, data from the *Perceptions of Practices Survey* can be used to determine the extent to which professional learning initiatives translate to practice. Through assessment using the *Survey*, leaders facilitating professional learning can determine which aspects of the problem-solving process are perceived as occurring more or less in schools. After establishing a baseline, schools and/or districts can adapt professional development plans to focus on specific steps of the problem-solving process that data indicate are a greater area of perceived need. Then, after providing targeted professional development, trainers can assess the level to which the trainings are translating to perceived practices as measured by the *Survey*. Additionally, MTSS leadership teams could share the data with educators in their school to increase educators' awareness and knowledge of the components of problem-solving that are strongly implemented and those that are in need of improvement. Recent literature on professional development indicates that learners should be actively involved in their learning and evaluate their personal growth to increase their ownership of the process (Haslam, 2010; Learning Forward, 2011).

One mechanism to activity engage educators in the process is for MTSS leadership teams to create Professional Learning Communities (PLC; Learning Forward, 2011) where educators analyze and discuss their data from the survey. In the PLC, educators can begin by examining their baseline data from the survey. The analysis of the data could include looking at each factor to determine general areas of problem-solving where improvement is needed or analyzing individual item responses. After exploring the data, the PLC could set goals they would like to achieve. Throughout the year, the team could continuously collect survey data to monitor and evaluate their progress toward the established goals. Through this process, educators may develop more buy-in and understanding of the components of problem-solving that are in need of improvement. Moreover, the team's on-going collection, analyzing, and use of data may encourage motivation, which researchers suggest will increase the likelihood of continuous improvement (Learning Forward, 2011).

Conclusion

MTSS implementation efforts are increasing in schools across the nation. The *Perceptions of Practices Survey* is a self-report tool that measures educators' perceptions of MTSS practices with a focus on problem-solving. The current study provided evidence for the reliability and validity of the measure using multilevel statistical techniques that account for nested data. Results of these analyses indicated a four-factor structure that provides data on problem-solving at all three tiers. In addition, significant relationships were found among the *Perceptions of Practices Survey* and the *Tiers I & II Critical Components Checklist* for the 2010 sample providing some evidence of concurrent validity. Thus, the *Perceptions of Practices Survey* is one available tool for educators to formatively evaluate and measure the fidelity of their MTSS implementation.

REFERENCES

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317-332.

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

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York, NY: Routledge.

Batsche, G. M., Curtis, M. J., Dorman, C., Castillo, J. M., & Porter, L. J. (2007). The Florida problem-solving/response to intervention model: Implementing a statewide initiative. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 378-395). New York, NY: Springer.

Batsche, G. M., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.

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

Bergan, J. & Kratochwill, T. R. (1990). *Behavioral Consultation and Therapy*. New York: Plenum Press.

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*, 23-33.

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, (42)*1, 85-95.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco: Jossey-Bass.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* New York, NY: Guilford Press.

Brown-Chidsey, R. & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice* (2nd ed.). New York, NY: The Guilford Press.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394. doi:10.1177/073428290502300406

Burns, M. K., & VanDerHeyden, A. M. (Eds). (2006). Using Response to Intervention to assess learning disabilities [Special issue]. *Assessment for Effective Intervention, 32*(1).

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

Castillo, J. M., Batsche, G. M., Curtis, M. J., Stockslager, K., March, A., Minch, D., & Hines, C. (2010). *Problem-solving/response to intervention evaluation tool technical assistance manual.* Retrieved from http://www.floridarti.usf.edu/resources/program_evaluation/ta_manual_revised2013/ta_manual_revised2013.pdf

Castillo, J. M., & Curtis, M. J. (2014). Best practices in systems-level change. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Systems-level services* (pp. 11–28). Bethesda, MD: National Association of School Psychologists.

Castillo, J. M., Dedrick. R. F., March, A. L., Stockslager, K. M., Hines, C. V., & Tan, S. Y. (2015). Development and initial validation of a scale measuring the beliefs of educators regarding Response to Intervention. *Journal of Applied School Psychology, 31,* 1-30.

Castillo, J. M., Hines, C. V., Batsche, G. M., & Curtis, M. J. (2011). *The Florida problem-solving/response to intervention project: Year 3 evaluation report.* Retrieved from http://www.floridarti.usf.edu/resources/format/pdf/yr3_eval_report.pdf

Center on Response to Intervention (2014). *RTI Fidelity of Implementation Rubric.* Washington, D.C.: American Institutes for Research.

Cronbach, L. J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, D.C.: American Council on Education.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders, 19*(1), 27-40. doi: 10.1177/1063426610365879

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

D'Haenens, E., Van Damme, J., & Onghena, P. (2012). Constructing measures for school process variables: the potential of multilevel confirmatory factor analysis. *Quality and Quantity, 46,* 155-188.

Donovan, M. S., & Cross, C. T. (Eds.) (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Duhon, G. J., Mesner, E. M., Gregerson, L., & Witt, J. C. (2009). Effects of public feedback during response to intervention team meetings on teacher implementation integrity and student academic performance. *Journal of School Psychology, 47*, 19 – 37.

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly, 16,* 149-167.

Erickson, A. G., Noonan, P. M., & Jenson, R. (2012). The School Implementation Scale: Measuring Implementation in Response to Intervention Models. *Learning Disabilities—A Contemporary Journal*, *10*, 33-52.

Fletcher, J. M., Francis, D. J., Morris, R. D., & Lyon, G. R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 506 – 522.

Fletcher, J. M., & Vaughn, S. (2009). Response to Intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30-37.

Florida's Multi-Tiered System of Supports (MTSS; 2014). *Foundation Beliefs.* Retrieved from http://www.florida-rti.org/oneSystem/beliefs.htm

Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Thousand Oaks, CA: Corwin.

Gansle, K. A., & Noell, G. H. (2007). The fundamental role of intervention implementation in assessing response to intervention. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden. (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 244 – 251). New York: Springer.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*, 72-91.

Gettinger, M., & Stoiber, K. (2007). Applying a response-to-intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education, 27*, 198–213.

Glover, T. A., & DiPerna, J. C. (2007). Service delivery for Response to Intervention: Core components and directions for future research. *School Psychology Review, 36*(4), 526-540.

Griffiths, A., Parson, L. B., Burns, M. K., VanDerHeyden, A. M., & Tilly, W. D. (2007). *Response to Intervention: Research for practice*. Alexandria, VA: National Association of State Directors of Special Education, Inc.

Gutkin, T. B., & Curtis, M. J. (2009). School-based consultation: The science and practice of indirect service delivery. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 591–635). Hoboken, NJ: John Wiley.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (1998) *Multivariate data analysis* (5th ed.) Prentice-Hall:London.

Hall, G. E., & Hord, S. M. (2015). *Implementing change: Patterns, principles and potholes* (4th ed.). Boston, MA: Pearson.

Haslam, M. B. (2010, January). *Teacher professional development evaluation guide*. Oxford, OH: National Skill Development Corporation.

Hauerwas, L. B., Brown, R., & Scott, A. N. (2013). Specific learning disability and response to intervention: State-level guidance. *Exceptional Children, 80*, 101–120. doi:10.1177/0014402913080001

Heller, K. A., Holtzman, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, DC: National Academy Press.

Hoover, J. J., Baca, L., Wexler-Love, E., & Saenz, L. (2008). *National implementation of response to intervention (RTI): Research summary*. Boulder: University of Colorado, Special Education Leadership and Quality Teacher Initiative, BUENO Center-School of Education. Retrieved from www.nasdse.org/Portals/0/NationalImplementationofRTI/ResearchSummary.pdf

Hosp, J. L., & Reschly, D. J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education, 37*(2),67-80. doi:10.1177/00224669030370020201

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

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

Hughes, C., & Dexter, D. D. (2011). *Field studies of RTI programs, Revised.* Retrieved from http://www.rtinetwork.org/learn/research/field-studies-rti-programs

Individuals with Disabilities Education Improvement Act, U.S.C. H.R. 1350 (2004).

Jackson, S. (2009). *Research methods and statistics: A critical thinking approach.* Belmont, CA: Wadsworth, Cenage Learning.

James Bell Associates. (2009). *Evaluation brief: Measuring implementation fidelity*. Retrieved from https://jbassoc-public.sharepoint.com/ReportsPublications/Evaluation%20Brief%20-%20Measuring%20Implementation%20Fidelity_Octob%E2%80%A6.pdf

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

Kelleher, C., Riley-Tillman, T. C.,&Power, T. J. (2008). An initial comparison of collaborative and expert-driven consultation on treatment integrity. *Journal of Educational and Psychological Consultation, 18*, 294 – 324.

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology In The Schools*, *49*, 342-352. doi:10.1002/pits.21602

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Kristapovich, P., Notter, L., & Robles-Villalba, V. (2014). *The Condition of Education 2014* (NCES 2014-083). T. Nachazel & A. Dziuba (Eds.). Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Kincaid, D., Childs, K., & George, H. (2010). *School-Wide Benchmarks of Quality – Revised* (Unpublished instrument). University of South Florida, Tampa.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco, CA:Jossey-Bass Inc.

Kratochwill, T. R., Volpiansky, P., Clements, M., & Ball, C. (2007). Professional development in implementing and sustaining multitier prevention models: Implications for Response to Intervention. *School Psychology Review, 36*(4), 618-631.

Learning Forward. (2011). *Standards for professional learning*. Oxford, OH: Author.

Lerner, J. W. (2002). *Learning disabilities: Theories, diagnosis, and teaching strategies* (8th ed.). Boston: Houghton-Mifflin.

Lyon, G. R, Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgeson, J. K., Wood, F. B., …& Olson, R. (2001). Rethinking learning disabilities. In C. E. Finn, Rotherham, A. J., & Hokanson, C. R. (Eds.), *Rethinking special education for a new century* (pp. 259–288). Washington, DC: Thomas B. Fordham Foundation and the Progressive Policy Institute.

Mellard, D. F., Frey, B. B., & Woods, K. L. (2012). School-wide student outcomes of response to intervention frameworks. *Learning Disabilities: A Contemporary Journal, 10*(2), 17-32.

Mislevy, J. L., & Rupp, A. A. (2010). Concurrent validity. In Salkind, N. J. (Ed.), *Encyclopedia of research design*. Los Angeles, CA: SAGE. doi: http://dx.doi.org/10.4135/9781412961288.n67

Mowbray, C. T., Bybee, D., Holter, M., & Lewandowski, L. (2006). Validation of a fidelity rating instrument for consumer-operated services. *American Journal of Evaluation, 27*, 9-27.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*(3), 315-340.

Murray, C. S., Woodruff, A. L., & Vaughn, S. (2010). First-grade student retention within a 3-tier reading framework. *Reading and Writing Quarterly, 26,* 26–50.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376–398.

Muthén, L.K., & Muthén, B.O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Muth´en, L. K., & Muth´en, B. O. (n.d.). *Multilevel modeling with latent variables using Mplus*. Unpublished manuscript.

No Child Left Behind Act, U.S.C. 115 STAT. 1426 (2002).

Noell, G. H., & Gansle, K. A. (2006). Assuring the form has substance: Treatment plan implementation as the foundation of assessing responses to intervention. *Assessment for Effective Intervention, 32*, 32-39.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Resetar, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*, 87-106.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

O'Connell, A., & McCoach, D.(Eds.). (2008). *Multilevel modeling of educational data*. Charlotte, NC: Information Age.

O'Conner, E. P., & Freeman, E. W. (2012). District-level considerations in supporting and sustaining RTI implementation. *Psychology in the Schools, 49*(3), 297-310. doi: 10.1002/pits.21598

Pluymert, K. (2014). Problem-solving foundations for school psychological services. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 25–40). Bethesda, MD: National Association of School Psychologists.

Prasse, D. P. (2014). Why adopt an RTI model? *RTI Action Network.* New York, NY: The National Center for Learning Disabilities, Inc.

President's Commission on Excellence in Special Education (PCESE; 2002). *A new era: Revitalizing special education for children and their families* (U.S. Department of Education Contract No. ED-02-PO-0791). Washington, DC: U.S. Department of Education.

Raffalovich, L. E., Deane, G. D., Armstrong, D., & Hui-Shien, T. (2008). Model selection procedures in social research: Monte-Carlo simulation results. *Journal Of Applied Statistics, 35*, 1093-1114. doi:10.1080/03081070802203959

Raudenbush, S. & Bryk, T. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). New York, NY: SAGE.

Raudenbush, S. W., Rowan, B., & Jin Kang, S. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high school data. *Journal of Educational and Behavioral Statistics, 16*, 295–330. doi:10.3102/10769986016004295

Reschly, D. J. (2008). School psychology paradigm shift and beyond. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (3-15). Bethesda, MD: National Association of School Psychologists.

Roach, A.T., Lawton, K., and Elliott, S.N. (2014). “Best practices in facilitating and evaluating the integrity of school-based interventions.” In A. Thomas and P. Harrison (Eds.) *Best Practices in School Psychology: Data-Based and Collaborative Decision-Making*. (pp. 133-146). Bethesda, MD: National Association of School Psychologists.

Sanetti, L. l., & Collier-Meek, M. (2014). Increasing the Rigor of Procedural Fidelity Assessment: An Empirical Comparison of Direct Observation and Permanent Product Review Methods. *Journal Of Behavioral Education*, *23*, 60-88. doi: 10.1007/s10864-013-9179-z

Sansosti, F. J., Noltemeyer, A., & Goss, S. (2010). Principals’ perceptions of the importance and availability of Response to Intervention practices within high school settings. *School Psychology Review, 39*, 286–295.

Schmoker, M. (2006). *Results now.* Alexandria, VA: Association for Supervision and Curriculum Development.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Spectrum K12 School Solutions. (2011). *Response to intervention adoption survey.* Retrieved from http://www.spectrumk12.com/resources/white_papers/response_to_intervention_adoption_survey

Snijders, T. (2005). Power and sample size in multilevel modeling. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1570–1573). Hoboken, NJ: Wiley.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically-based tests for the number of common factors*. Paper presented at the annual meeting of the Psychonomic Society, Iowa City, IA.

Stoiber, K. C. (2014). A comprehensive framework for multitiered systems of support in school psychology. In P. Harrison & A. Thomas (Eds.), *Best Practices in School Psychology VI* (41-70). Bethesda, MD: National Association of School Psychologists.

Sugai, G., Horner, R., Lewis-Palmer, T., & Rossetto Dickey, C. (2012). *Team Implementation Checklist, Version 3.1 (TIC v. 3.1).* Eugene, OR: University of Oregon.

Sugai, G., Horner, R.H., & Todd, A.W. (2003). *Effective Behavior Support (EBS) Survey: Assessing and planning behavior support in schools.* Eugene, OR: University of Oregon.

Sugai, G., Lewis-Palmer, T., Todd, A.W., & Horner, R.H. (2001). *School-Wide Evaluation Tool (SET)*. Eugene, OR: University of Oregon.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC, US: American Psychological Association.

Torgeson, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in Reading First schools. *Child Development Perspectives, 3*, 38-40. doi:10.1111/j.1750-8606.2009.00073.x

United States Department of Education (2012). *What ESEA flexibility means for students, teachers, and parents: Answering the public's questions.* Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education, Office of Planning, Evaluation and Policy Development (2010). *ESEA Blueprint for Reform*, Washington, D.C.

VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of a response to intervention (RTI) model on identification of children for special education. *Journal of School Psychology, 45*, 225-256. doi:10.1016/j.jsp.2006.11.004

Vaughn, S., & Fuchs, L. S. (2006). A response to "competing views: A dialogue on Response to Intervention": Why Response to Intervention is necessary but not sufficient for identifying students with learning disabilities. *Assessment for Effective Intervention, 32*(1), 58-61.

Zirkel, P. A., & Thomas, A. B. (2010). State laws for RTI: An updated snapshot. *Teaching Exceptional Children, 42*(3), 56-63.

APPENDIX A:

Perceptions of Practices Survey

Perceptions of Practices Survey

1. **Your PS/RtI Project ID:**

   Your PS/RtI Project ID was designed to assure confidentiality while also providing a method to match an individual's responses across instruments. In the space provided (first row), please write in the last four digits of your Social Security Number and the last two digits of the year you were born. Then, shade in the corresponding circles

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

Directions: *For each item on this survey, please indicate how frequently or infrequently the given practice occurs in your school for both academics (i.e., reading and math) and behavior. Please use the following response scale*:

1 = **Never Occurs *(NO)***

2 = **Rarely Occurs *(RO)***

3 = **Sometimes Occurs *(SO)***

4 = **Often Occurs *(OO)***

5 = **Always Occurs *(AO)***

¡ = **Do Not Know (DK)**

In my School:	NO	RO	SO	OO	AO	DK
2. Data (e.g., Curriculum-Based Measurement, DIBELS, FCAT, Office Discipline Referrals) are used to determine the percent of students receiving core instruction (general education classroom only) who achieve benchmarks (district grade-level standards) in:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
3. Data are used to make decisions about necessary changes to the core curriculum or discipline procedures to increase the percent of students achieving benchmarks (district grade-level standards) in:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
4. Data are used (e.g., Curriculum-Based Measurement, DIBELS, Office Discipline Referrals) to identify at-risk students in need of supplemental and/or intensive interventions for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
5. The students identified as at-risk routinely receive additional (i.e., supplemental) intervention(s) for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡

In my School:	N O	R O	S O	O O	A O	D K
6. Progress monitoring occurs for all students receiving supplemental and/or intensive interventions for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
7. Progress monitoring data (e.g., Curriculum-Based Measurement, DIBELS, behavioral observations) are used to determine the percent of students who receive supplemental and/or intensive interventions who achieve grade-level benchmarks for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
8. A standard protocol intervention (i.e., the same type of intervention used for similar problems) is used initially for all students who require supplemental instruction for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡

***Directions:** Items 9-18 refer to the typical Problem-Solving Team (i.e., Student Support Team, Intervention Assistance Team, School-Based Intervention Team, Child Study Team) meeting in your school that includes a student who has been referred for problem-solving or a special education evaluation. While addressing each item for academics (math and reading), think of a typical case in which a student has been referred for an academic concern. While addressing each question for behavior, think of a typical case in which a student has been referred for a behavioral concern. Then, please indicate how frequently each of the given practices occurs in your school using the same scale.*

In my School:	NO	RO	SO	OO	AO	DK
9. The target behavior is routinely defined in terms of the <u>desired</u> behavior (e.g., Johnny will raise his hand to ask a question, Susie will read 90 correct words per minute) instead of the <u>problem</u> behavior (e.g., Johnny talks out of turn, Susie reads below grade-level) for:						
c. Academics	1	2	3	4	5	¡
d. Behavior	1	2	3	4	5	¡
10. Quantifiable data (e.g., reading fluency score, percent compliance, percent on-task behavior) are used to						
a. identify the target student's current performance in the area of concern for:						
• Academics	1	2	3	4	5	¡
• Behavior	1	2	3	4	5	¡
b. identify the <u>desired</u> level of performance (i.e., the benchmark) in the area of concern for:						
• Academics	1	2	3	4	5	¡
• Behavior	1	2	3	4	5	¡
c. identify the current performance of same-age peers using the same data as the target student for:						
• Academics	1	2	3	4	5	¡
• Behavior	1	2	3	4	5	¡
11. The Problem-Solving Team routinely develops hypotheses (i.e., proposed reasons) explaining why the target student is not demonstrating the <u>desired</u> behavior for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡

In my School:	NO	RO	SO	OO	AO	DK
12. Data are collected to confirm the reasons that the student is not achieving the desired level of performance for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
13. Intervention plans are routinely developed based on the confirmed reasons that the student is not achieving the desired level of performance for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
14. The teacher of a student referred for problem-solving routinely receives staff support to implement the intervention plan developed by the Problem Solving Team for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
15. Data are collected routinely to determine the degree to which the intervention plans are being implemented as intended for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
16. Data are graphed routinely to simplify interpretation of student performance for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡
17. Progress monitoring data are used to determine						
a. the degree to which the target student's rate of progress has improved for:						
• Academics	1	2	3	4	5	¡

In my School:	NO	RO	SO	OO	AO	DK
• Behavior	1	2	3	4	5	¡
b. whether the gap has decreased between the target student's current performance and the desired level of performance (i.e., benchmark) for:						
• Academics	1	2	3	4	5	¡
• Behavior	1	2	3	4	5	¡
c. whether the gap has decreased between the target student's current performance and the performance of same-age peers for:						
• Academics	1	2	3	4	5	¡
• Behavior	1	2	3	4	5	¡
18. A student's response-to-intervention data (e.g., rate of improvement) are used routinely to determine whether a student is simply behind and <u>can</u> learn new skills <u>or</u> whether the student's performance is due to a disability for:						
a. Academics	1	2	3	4	5	¡
b. Behavior	1	2	3	4	5	¡

THANK YOU!

APPENDIX B:

Tiers I and II Critical Components Checklist

Tiers I and II Critical Components Checklist

School: ______________________ **Target Area:** ☐ Reading ☐ Math ☐ Behavior

Window: ☐ 1 ☐ 2 ☐ 3 **Grade Level (if applicable):** ______________

<u>Directions</u>: For each selected target area and grade-level, please use the scale provided to indicate the degree to which each critical component of a Problem-Solving/Response to Intervention (PS/RtI) model is present in paperwork (i.e., permanent products) derived from data meetings (i.e., meetings in which the PS/RtI model is used to examine Tier I and/or II instruction). See the attached rubric for the criteria for determining the degree to which each critical component is present in the paperwork.

Component	0 = Absent 1 = Partially Present 2 = Present N/A = Not Applicable	Evidence/Comments
Problem Identification		
1. Data were used to determine the effectiveness of core instruction	0 1 2	
2. Decisions were made to modify core instruction or to develop supplemental (Tier II) interventions	0 1 2	
3. Universal screening (e.g., DIBELS, ODRs) or other data sources (e.g., district-wide assessments) were used to identify groups of students in need of supplemental intervention	0 1 2	
Problem Analysis		
4. The school-based team generated hypotheses to identify potential reasons for students not meeting benchmarks	0 1 2	
5. Data were used to determine viable or active hypotheses for why students were not attaining benchmarks	0 1 2	
Intervention Development and Implementation		
6. Modifications were made to core instruction		
a. A plan for implementation of modifications to core instruction was documented	0 1 2 N/A	
b. Support for implementation of modifications to core instruction was documented	0 1 2 N/A	
c. Documentation of implementation of modifications to core instruction was provided	0 1 2 N/A	

Component	0 = Absent 1 = Partially Present 2 = Present N/A = Not Applicable	Evidence/Comments
7. Supplemental (Tier II) instruction was developed or modified		
a. A plan for implementation of supplemental instruction was documented	0 1 2 N/A	
b. Support for implementation of supplemental instruction was documented	0 1 2 N/A	
c. Documentation of implementation of supplemental instruction was provided	0 1 2 N/A	
Program Evaluation/RtI		
8. Criteria for positive response to intervention were defined	0 1 2	
9. Progress monitoring and/or universal screening data were collected/scheduled	0 1 2	
10. A decision regarding student RtI was documented	0 1 2	
11. A plan for continuing, modifying, or terminating the intervention plan was provided	0 1 2	

Additional Comments: