8-2017

An Investigation of Multi-Tiered System of Supports: Implementation Perceptions and Third Grade Reading Achievement

Valerie JH Sherman
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

AN INVESTIGATION OF MULTI-TIERED SYSTEM OF SUPPORTS: IMPLEMENTATION PERCEPTIONS AND THIRD GRADE READING ACHIEVEMENT

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Valerie JH Sherman

College of Education and Behavioral Sciences
School of Special Education

August 2017

This Dissertation by: Valerie JH Sherman

Entitled: *An Investigation of Multi-Tiered System of Supports: Implementation Perceptions and Third Grade Reading Achievement*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in the College of Education and Behavioral Sciences in School of Special Education.

Accepted by the Doctoral Committee

____

Corey D. Pierce, Ph.D., Research Advisor

____

John Luckner, Ed.D., Committee Member

____

Jennifer Urbach, Ph.D., Committee Member

____

Donna Bottenberg, Ph.D., Faculty Representative

Date of Dissertation Defense ___________________________________

Accepted by the Graduate School

____

Linda L. Black, Ed.D.
Associate Provost and Dean
Graduate School and International Admissions

ABSTRACT

Sherman, Valerie JH. *An Investigation of Multi-Tiered System of Supports: Implementation Perceptions and Third Grade Reading Achievement*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2017.

The multi-tiered system of supports (MTSS) is intended to provide ongoing support and needs-based professional development for teachers who are (a) designing and delivering instruction, (b) administering universal screeners to identify students who are at risk, and (c) using the data those screeners generate during their instructional planning process and while making placement decisions. However, there is a lack of national consensus on the critical components of the MTSS framework, how those components should be defined, and whether individual elements have a greater impact on student reading outcomes than others. While many noted the MTSS initiative has the potential to positively impact student outcomes, research also demonstrated professional educators struggle to implement the model effectively. If the MTSS initiative is to survive deep into the 21st century, research must demonstrate it has the potential to positively impact student reading achievement, and help clarify the essential components for those vested in the implementation. The primary purpose of this study was to investigate how educator perceptions of MTSS implementation in Colorado (*n* = 376) related to the reading outcomes of elementary students. A secondary purpose sought to identify the individual components of the MTSS framework currently in use within Colorado to discern if individual factors of the MTSS framework impacted student reading outcomes more than others.

Structural equation modeling (SEM) was used to test this study's hypothesized models; when viewed comprehensively, the results indicated when an MTSS framework included components associated with (a) leadership, (b) evidence-based instructional practices, (c) universal screening and progress monitoring, (d) data-based problem solving, and (e) partnerships between families and schools, student reading outcomes tended to improve. Implications of the study indicated the MTSS has the potential to counteract an important portion of the impact poverty has on the reading outcomes of students who struggle while learning to read and is an effective system that can be used by educators to have a meaningful and long-term impact on their students, their communities, and the nation at large.

ACKNOWLEDGEMENTS

There is an African proverb that wisely notes it takes a village to raise a child. At this point in my doctoral journey of learning, I realize that it also takes a village to raise a doctoral learner. There are many people I would like to thank for their support and encouragement during my doctoral program.

First, Dr. Corey Pierce, you have been an amazing mentor and advocate. You helped pave my doctoral path with both learning and laughter. You inspired me to give my all, do my best, and believe in myself. You have been amazing. Thank-you for letting me add my name to this list of those whom you have advised.

Drs. Luckner, Urbach and Bottenberg, you have all helped shape my thinking, taught me how to teach, achieve, and have been a stable source of encouragement and support. I thank you each of you for the time you have devoted to helping me learn and grow. I am a better person, scholar, and researcher because of it.

I also want to thank Dr. Nancy Sileo, Dr. Eugene Sheehan, and everyone in the College of Education and Behavioral Sciences. Being a doctoral research assistant for the College helped me grasp what being a member of higher education could look like if you worked hard enough and asked the right question.

I also want to thank my fellow doctoral learners. Without friends like Mary Anne, Raveema, Jean, and Marlo, my learning and growth would have been curtailed. You are all amazing women!

Thanks also to Dr. Susan Hutchinson, structural equation modeling researcher and professor extraordinaire. Without your advice, expertise, and knowledge, this project would have been insurmountable. Thank you, from the bottom of my heart, for everything.

I would also like to thank Dr. Scott Ross and his colleagues from the Office of Learning and Supports within the Colorado Department of Education. Because of you, I know more about MTSS, can proudly say my dissertation was funded by a grant, and have a deep appreciation for how hard you work to support Colorado educators.

Finally, I want to thank the members of my family because without your support, love, understanding, and well-developed abilities to both leave me alone when I needed to work and welcome me back into the fold with open arms when I needed a break, this would not have been possible. Scott, you are an amazing husband, friend, and soldier. Our dads sure knew what they were doing when they introduced us. Thanks for letting me be a part of your journey and for being a part of mine. Trevor, you are simply remarkable. Your kind heart, firm convictions, and cunning humor make my life better. You know what to say and what to do to help make a bad day better and turn a good day into a great one. I am one lucky mom to have you as a son. Natalie, you are a wonder. Your courageous spirit, deep intellect, and desire to make the world a better place inspires me. I cannot wait to see what the future holds for you. Thanks for letting me be a part of your life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO THE STUDY

Reading is an essential skill students need to master during their education in order to learn about the world they live in, communicate with others effectively, maximize their individual potential, and lead fulfilling lives. Historically, researchers have found when students do not learn to proficiently read early during their education, they learn less than their peers (Cunningham & Stanovich, 1997; Ensminger & Slusarcick, 1992; Juel, 1988), have lower levels of self-esteem (Rose, 2006), are more likely to drop out of school before graduating (Bost & Riccomini, 2006; Compton et al., 2012; Ensminger & Slusarcick, 1992), tend to demonstrate problematic behavior more frequently (McIntosh, Horner, Chard, Dickey, & Braun, 2008), and suffer socially (Brynner, 2008). These findings are concerning on their own but become more alarming when one considers recent National Assessment of Educational Progress data (cited in Kena et al., 2015), which revealed only one-third of all students in the United States are able to read at or above the proficient level while the remaining two-thirds struggle.

Moats (2009) shared that as a result of this scholastic melt-down, 21st century educational policy reforms and federal legislative initiatives have focused on improving the reading outcomes of students who struggle and sought to identify the instructional strategies classroom educators should utilize to remediate those difficulties. For example, Reading First was part of the No Child Left Behind Act of 2001 (NCLB) and provided federal funding to train teachers and assist the striving readers those teachers served (Torgesen, 2009). Additionally, funding provided by the National Institute of Child Health and Human Development was used by the National Reading Panel (NRP; 2000) to identify and describe five essential components of successful reading instruction.

The NRP report had a far-reaching impact on reading instruction because the authors used the body of research to definitively identify the skills students must master to become proficient readers, which, in turn, provided an instructional focus for the teachers serving those students (Moats, 2009). Those components are (a) phonemic awareness--the knowledge that spoken words are made up smaller segments of sound or phonemes; (b) phonics--the understanding that written letters and groups of letters are used to represent sounds and that those letters can be combined to represent words; (c) fluency--the skill readers use to recognize words easily, read with greater speed, accuracy, and expression, and understand what is being read; (d) vocabulary--knowing the meaning of words; and (e) comprehension--the act of understanding the information presented in a text (NRP, 2000).

Unfortunately, despite the attention that reading policies and instructional practices have garnered since 2000, little practical progress with students has been made (Kena et al., 2015). While research findings consistently demonstrate every student is capable of reading either at or above grade-level by the end of first grade (e.g., Denton et al., 2011; Mathes, Denton, Anthony, Francis, & Schatschneider, 2005; Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008; Vellutino et al., 1996), many students continue to struggle while learning to read, fall further and further behind their grade-level peers, and eventually require the costly educational services and supports provided by an individualized education plan (IEP; Stanovich, 1986).

Response to Intervention

In response to the rising numbers of students qualifying for an IEP, federal policy makers who drafted the most recent reauthorization of the Individuals with Disabilities Education Act (IDEA; U.S. Department of Education, 2004) provided an alternative method to the intelligent quotient discrepancy testing and identification model, which was the primary method school-based professionals had been using to qualify students in the area of specific learning disabilities for the services of an IEP. This alternative method used data from standardized curriculum-based assessments and progress monitoring measures to identify students who persistently struggled and provided them with supplemental and increasingly intensive instructional interventions and educational supports in small-group settings. During those interventions, student responses were monitored more frequently and those data were used to both guide the instructional planning process and make placement decisions. Originally, this alternative process, coined response-to-intervention (RTI), was simply intended to more accurately identify specific students who required individualized special educational services provided by an IEP (Johnston, 2010; Kame'enui, 2007; Shinn, 2007; Zirkel, 2011). Today, in addition to improving the accuracy of special education identification, RTI is a driving force in general educational reform initiatives (Fuchs & Vaughn, 2012).

From those early beginnings, RTI has evolved into a general education intervention model used to provide all students with differentiated, evidence-based instruction that is paired with supplemental and increasingly intensive (tiered) interventions for students who struggle to meet grade-level expectations (Bursuck & Blanks, 2010). Students who receive those tiered interventions are regularly monitored for progress to determine if their responses would enable them to catch up to their grade-level peers in a timely manner (Gehsmann, 2008). According to Fletcher and Vaughn (2009), one of the primary purposes of current-day RTI is to provide students with tiered interventions that become more intensive when students fail to respond to the universal instruction offered within Tier I settings to decrease the probability they develop long-term academic difficulties that become more difficult to correct over time.

A variety of researchers have studied how the implementation of the RTI instructional framework affects student reading achievement (Bursuck & Blanks, 2010) and noted how the increasingly intensive tiered levels of support have the potential to positively impact student reading outcomes (e.g., Al Otaiba, Kim, Wanzek, Petscher, & Wagner, 2014; Compton et al., 2012; Miller et al., 2014). However, the RTI framework also has its critics (Balu, Zhu, Doolittle, Schiller, & Jenkins, 2015). These opponents justifiably stated the language used in IDEA (U.S. Department of Education, 2004) did not provide any detailed recommendations about the individual components of the RTI model and failed to offer specific implementation guidelines educators and administrators could use during real-world scale-up efforts. Additionally, clear definitions of the terms interventions, responsiveness, and non-responsiveness were all left to be operationalized by educators charged with the important task of making a difference in the lives of their students. However, experts largely agreed that adopting universal screening assessment processes, monitoring the progress of students who need more intensive supports, and using data to guide the instructional planning process and make placement decisions should be included in any RTI scale-up effort (Gersten et al., 2009; Mesmer & Mesmer, 2008). Regardless of the difficulties presented by its lack of clarity, RTI is viewed by many as a tool that can be used by teachers and administrators to increase student learning. As a result, many states and school-districts are working to incorporate the tiered instructional supports of the RTI model into their local educational blueprints (Hughes & Dexter, 2011).

Positive Behavior Intervention and Supports

Like RTI, school-wide positive behavioral interventions and supports (SW-PBIS) is a universal, school-wide prevention strategy being used in schools across the United States to improve student learning. Specifically, the SW-PBIS framework was designed to positively modify school and district environments by using policies, systems, and practices to stimulate positive behavioral change for students, teachers, and administrators alike (Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008). According to Bradshaw, Koth, Bevans, Ialongo, and Leaf (2008), the ultimate goal of the SW-PBIS framework is multi-faceted: SW-PBIS seeks to limit disruptive behaviors that negatively impact educational environments and simultaneously improve the overall organizational health of schools. As such, many of the SW-PBIS programs strive to systematically manage student behavior by creating school-wide plans that transparently define and describe behavior expectations, incentivize positive behavior, and utilize a uniform approach to address problematic behaviors (Sugai & Horner, 2006).

Unfortunately, LaVigna and Willis (2012) shared that experts who endorsed SW-PBIS have also struggled to create common-sense SW-PBIS pedagogical guidelines that can used by educators in real-world classroom settings. To demonstrate, consider a study conducted by Reinke, Stormont, Herman, Puri, and Goel (2011). Reinke et al. (2011) investigated the classroom management self-perceptions of 292 classroom teachers and reported the teachers who participated in their study (a) felt they continually struggled to positively manage student behavior, (b) indicated classroom management was the most difficult and challenging aspect of their job, and (c) believed they were provided with inadequate level of classroom management-related training and professional development. The body of research also demonstrated ineffective classroom behavior management practices were linked with negative outcomes for students and teachers alike. Students who were placed in classroom environments where behavior was ineffectively managed received smaller amounts of academic instruction and learned less than students in classrooms where the converse was true (e.g., Reinke, Herman, & Stormont, 2013). Additionally, teachers who had higher levels of stress as a result of problematic student behavior also had lower levels of self-efficacy than their non-stressed peers (Klassen & Chiu, 2010). However, the body of research also demonstrated that when SW-PBIS was brought to scale effectively, it had the potential to positively impact student learning (Lane, Menzies, Ennis, & Bezdek, 2013). As a result, local school districts, educational researchers, and policymakers continue to investigate, implement, and incorporate SW-PBIS into the local vernacular of school improvement efforts (e.g., Lane et al., 2013; Sugai & Horner, 2006).

The Multi-Tier System of Supports

Recently, a variety of states and school districts across America including Colorado, Florida, Kansas, Los Angeles and Boston formally recognized a link between academic achievement and behavior and are working to meld the student-centered academic supports of the RTI framework with the school-wide behavioral management system of SW-PBIS into a single framework (Sugai & Horner, 2009). This combined model is increasingly being referred to by educational researchers and policy makers as the multi-tiered system of supports or MTSS (Sugai & Horner, 2009). Like RTI and SW-PBIS, MTSS has the potential to improve long-term educational outcomes of all students regardless of ability level (Oakes, Lane, & Germer, 2014). The overarching purpose of the MTSS framework is to create sustainable systems-level change at both the classroom- and district-levels. In 2010, a practical description of MTSS was created by the Kansas Multi-Tier System of Supports team within the Kansas State Department of Education:

> The MTSS approach provides a framework to create a single system that has the availability of a continuum of multiple supports for all students…. When implemented fully, an effective MTSS results in a self-correcting feedback loop that uses universal screening assessment data to not only intervene at the student level, but also to continuously refine the system by analyzing grade, building, and district data for the purpose of school improvement. (p. 1)

In a recent review of state-wide MTSS-related systems, American Institutes for Research (cited in Bailey, 2017) scholars shared that 21 states have explicitly adopted a multi-tiered system of supports within their educational blueprints that integrates both academic and behavioral supports into a single system-level framework (see Figure 1). Many of these states continue to use the term RTI to describe the general educational framework, which is similar to MTSS. However, states with an MTSS framework are using RTI to describe their special education eligibility determination process, which creates a general level of confusion at the national-level (Bailey, 2017).

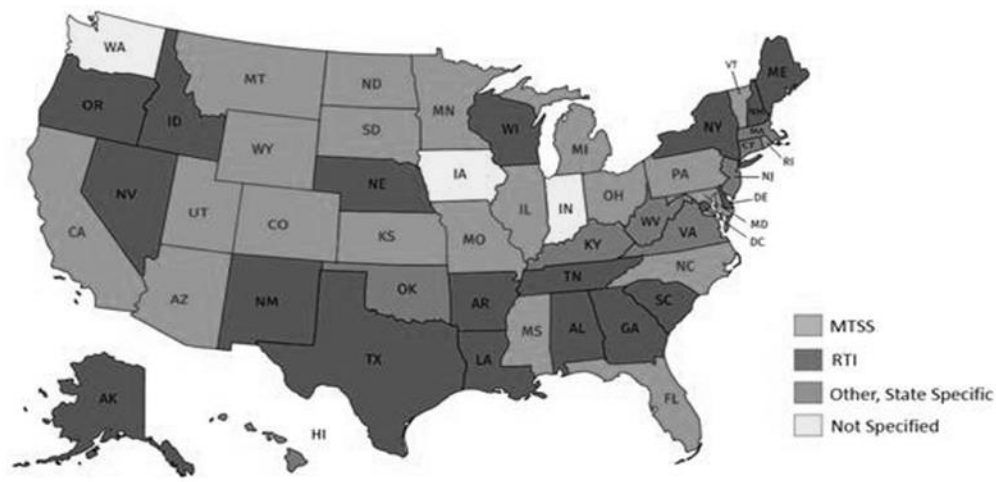


Figure 1. Multi-tiered system terminology by state.

According to Hurst (2014), the MTSS framework aligns resources and supports provided to both students and teachers while remaining focused on the scale-up and sustainability of school-wide improvement efforts. Expanding beyond the student-centered focus of the RTI framework, MTSS strives to ensure that instructional practices, educational policies, state and federal initiatives, and curricular programs are aligned at the classroom-, school- and district-levels (e.g., Harn, Chard, & Kame-enui, 2011; Lane et al., 2013; Utley & Obiakor, 2015; Vaughn et al., 2009). As such, MTSS is intended to provide ongoing support and needs-based professional development for teachers who are (a) designing and delivering instruction, (b) administering universal screeners to identify students who are at risk, and (c) using the data those screeners generate during their instructional planning process and while making placement decisions (Lane et al., 2013; Lane, Oakes, & Menzies, 2014; Sugai & Horner, 2009).

Statement of the Problem

As noted, there is a lack of national consensus on the critical components of the MTSS framework, how those components should be defined, and whether individual elements have a greater impact on student reading outcomes than others (Hudson, 2013; Samuels, 2016). The National Center on Response to Intervention (2012) identified (a) universal screening, (b) progress monitoring, (c) multi-level prevention, and (d) data-based decisions for their multi-tiered model. Many of the 21 states with a MTSS-type model made the decision to supplement the recommendations made by the National Center on Response to Intervention and add additional components to their individual frameworks. For example, 100% of the states included an evidence-based practices component, while components focused on shared leadership have been included by state-level leaders in Arizona, California, Colorado, Kansas, Oregon, Pennsylvania, Utah, and Virginia. Colorado, Florida, Michigan, Pennsylvania, Utah, and Virginia (or 29% of the total) stressed the importance of family, school, and community partnerships while14% (Arizona, California and Kansas) incorporated an integration and sustainability component. Other components not mentioned above included (a) classroom management (California); (b) early interventions and fidelity of implementation (Kansas and Michigan); (c) professional learning and support (Oregon, Pennsylvania, and Utah); (d) school culture (Oregon and Virginia); and (e) early identification (Pennsylvania). While this lack of component clarity at the state-level is the norm rather than the exception, individual districts and the teachers they employ continually strive to include the MTSS framework into their local educational blueprint with the hope improved student learning would follow.

Unfortunately, the evidence is mounting that individual schools, districts, and states continue to struggle during their scale-up and implementation efforts (e.g., Balu et al., 2015; Hudson, 2013). This might be because RTI and other similar multi-tiered initiatives require cooperation and collaboration at every level. In a 2016 Education Week interview with reporter Christina Samuels, national RTI expert and Vanderbilt professor Douglas Fuchs recently shared, “It would be unfair of anyone to come down on the schools in how they are implementing RTI because of the inherent complexity of the reform” (p. 2). Fuchs continued, sharing that individuals in many schools have diligently worked to make RTI work but continue to struggle because RTI resembles a complex machine with a wide array of working parts: “to get all those parts moving in synchrony is a very tall order” (p. 2). Similarly, Sherman (2016) found that while administrators and professional educators alike recognized the positive potential of the MTSS framework, they also struggled to understand the structural elements of a multi-tiered model, needed time to implement the initiative with fidelity, and indicated they craved both guidance and support. These findings supported those shared by Balu et al. (2015) who noted that schools struggle to implement MTSS with both accuracy and precision, which should not come as a surprise when the basic components of MTSS vary so widely by state. Additionally, while individual components of a MTSS framework vary between states, they can also vary within an individual state over time. For example, leaders in Colorado recently combined the universal screening and progress monitoring component with the problem-solving process component of their MTSS model, a change that has understandably led to increased levels of MTSS-related conceptual confusion from educators and administrators around Colorado.

Historically, many educational initiatives have faltered initially from a lack of conceptual clarity and technical adequacy during implementation and scale-up efforts (e.g., Moats, 2009; NRP, 2000, Zirkel, 2011). However, with time, commitment, and the occasional court ruling, those difficulties have typically been resolved (Turnbull, Turnbull, & Wehmeyer, 2010). It should come as no surprise that the novelty of the MTSS initiative means a coherent understanding of the framework might take some time to develop within America's schools. However, when student learning is at stake, time is a luxury the American educational system simply cannot afford. Fiester (2010) shared, "Low achievement in reading has important long-term consequences in terms of individual earning potential, global competitiveness, and general productivity" (p. 9). For example, the National Research Council (1998) shared that students whose reading proficiency was low had more behavioral and social problems than their peers, had lower levels of academic success in high school, and were less likely to graduate. The financial implications of failing to graduate over a lifetime are difficult to calculate but Planty, Hussar, and Snyder (2008) found individuals who did not have a high school diploma made $25,000 less per year than those who had graduated from college. This income discrepancy could lead to a separation between schools and families with low income levels, especially when the parents of students who struggle have low levels of education. Previous research demonstrated that all too often, schools tend to have low expectations of students from families with a lower socioeconomic status (SES; Fiester, 2010). In sum, when students fail to master the skill of reading within the first four years of their educational career, not only is it possible they will suffer but it is also possible the future generations of our society would be negatively affected. As aptly stated by former

American president Barack Obama, "The relative decline of American education is untenable for our economy, unstainable for our democracy, and unacceptable for our children, and we cannot afford to let it continue" (cited in Fiester, 2010, p. 4).

Purpose of the Study

While many noted the MTSS initiative has the potential to positively impact student outcomes, research also demonstrated that professional educators struggle to implement the model effectively. If the MTSS initiative is to survive deep into the 21st century, research must demonstrate it has the potential to positively impact student reading achievement and help clarify the essential components for those vested in the implementation. Therefore, this study had several purposes. The primary purpose of this study was to examine how perceptions of MTSS implementation in Colorado related to reading outcomes of elementary students. A secondary purpose sought to identify the individual components of the MTSS framework currently in use within Colorado to see if individual factors of the MTSS framework impacted student reading outcomes more than others.

Structural equation modeling (SEM) was used to test this study's hypothesized models (see Figures 2 and 3). Structural equation modeling is an ideal procedure for examining underlying theories of complex relationships among unobservable variables. Complex relationships are those with both direct and indirect effects of observable and unobservable variables.

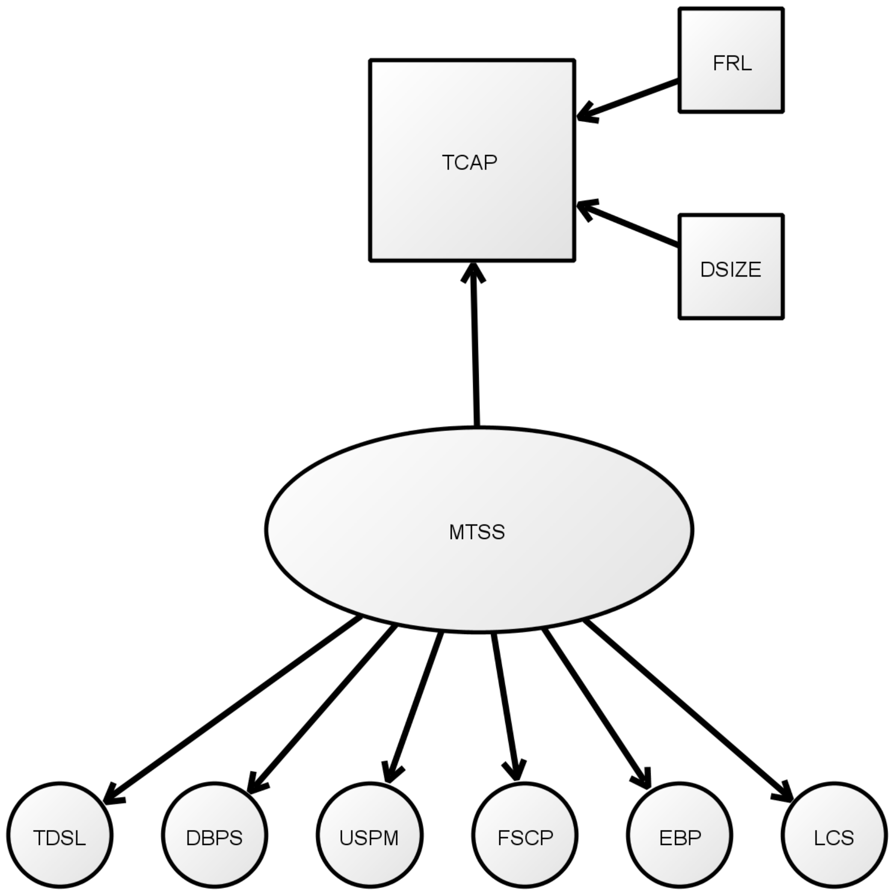


Figure 2. Hypothesized higher-order structural model.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; DSize= Standardized K-12 enrollment for each school's district; MTSS= Multi-Tiered System of Supports; TDSL= Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice; LCS= Layered Continuum of Supports.

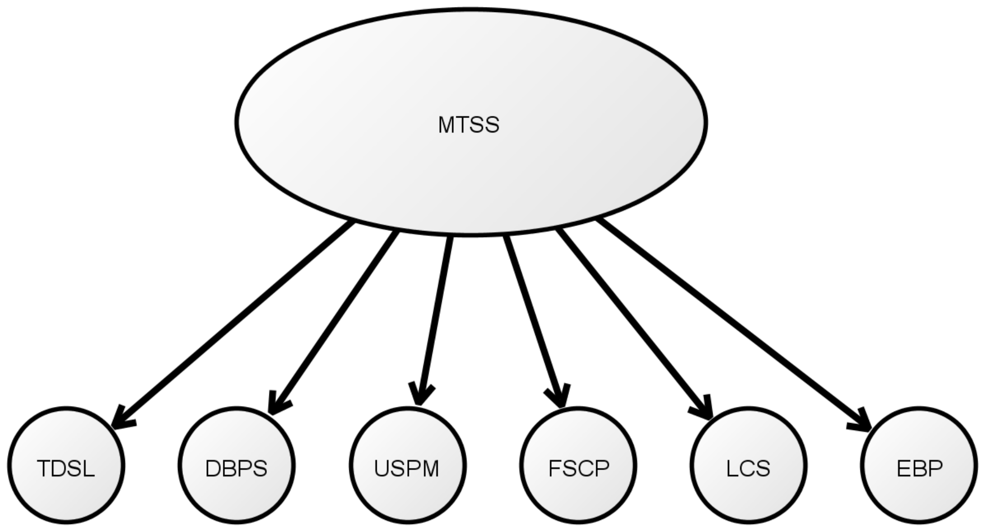


Figure 3. Higher-order confirmatory factor analysis model.
Note. MTSS= Multi-Tiered System of Supports; TDSL = Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM= Universal Screening and Progress Monitoring; FSCP= Family, School and Community Partnerships; LCS= Layered Continuum of Supports; EBP = Evidence-Based Practices.

Research Questions

This study examined the relationship between third grade student reading achievement and MTSS perceptions of implementation in Colorado by answering the following questions:

Q1 Does the hypothesized higher order MTSS theoretical factor structure of each measurement model fit the data?

Q1a For the proposed MTSS models hypothesizing relationships between implementation perception of MTSS and 2014 third grade Transitional Colorado Assessment Program (TCAP) data, do the data fit the models?

Q1b Does one model fit the data better than the others?

Q1c What effect does school-level percent of free and reduced lunch have on 2014 third grade TCAP reading scores?

Q1d What effect does district size have on TCAP scores?

Q1e Which latent factors account for more of the variance in student reading outcomes?

CHAPTER II

LITERATURE REVIEW

Studies included in this synthesis have been organized both by theme and in chronological order as frequently as possible to provide the reader with a clear sense of how the MTSS-related research has evolved over time. To increase the coherence of this synthesis, each theme begins with a general introduction to the MTSS-related concept. Summaries of individual studies and a brief theme-related discussion conclude each section. This synthesis ends with an overarching analysis of the general strengths of the research, how those studies contributed to the development of the MTSS framework, an analysis of the research designs used by the authors, and a general statement that describes the relevancy of the MTSS-related research.

The research process started with a database search by topic. Because MTSS is a topic uniquely educational, the search was conducted using only educational search engines. Education Source was used to identify a majority of the studies; however, ERIC and PsycInfo search engines also provided useful information. Because a wealth of research has been conducted that relates to multi-tiered systems of support and elementary reading skills, it quickly became apparent that using specific terms and search parameters would be necessary. To limit the scope of the search to the most useful and timely information, search terms primarily included (but were not limited to) RTI, elementary, reading, systems, experimental, and quasi-experimental. To report the most relevant and recent research, timeline parameters were limited to studies published from 2006 to the present.

The Impact of Technical Support, Communication, and Collaboration

Effective MTSS-scale up efforts all have one thing in common: the individuals involved in the process must provide teachers with timely and relevant assistance, communicate effectively, and work collaboratively with all the individuals involved in the process. The body of research that contributed to or confirmed the ideas supporting (a) the importance of providing just-in-time technical assistance and professional support, (b) effective communication, and (c) the impact of collaborative partnerships during successful multi-tiered systems scale-up efforts is detailed in the following narrative. Summaries of relevant experimental, quasi-experimental, and qualitative studies are included. At a minimum, the authors' purposes and findings are shared. In some instances, a brief description of the participants and settings is provided to create a more complete understanding of the research. This section concludes with a brief summary and analysis of the individual themes that connect these studies.

Technical Assistance, Professional Development, and Communication

A sample of studies noted the positive impact that providing teachers with technical assistance and professional support had on MTSS scale-up efforts. For example, a district-level profile by Gil and Woodruff (2011) qualitatively described the successful implementation of a multi-tiered system in a southern California district originally created to improve the literacy achievement of the district's English language learners. During the scale-up, both school- and district-level administrators provided timely technical assistance and relevant professional development opportunities to their teachers that appeared to positively impact student literacy outcomes. The authors also noted the scale-up of the intuitive was successful because classroom educators were also provided with sufficient time to (a) engage in purposeful conversations about expected outcomes, (b) build collaborative environments that facilitated professional discourse and personal learning, and (c) use data to provide students with the proper level of support.

In contrast, Regan, Berkeley, Hughes, and Brady (2015) examined the perceptions of 63 teachers who did not receive additional technical assistance or support in a northeastern school district that had recently adopted a multi-tiered system of supports using an electronic survey. Survey results revealed respondents felt well prepared to use both evidence-based instructional practices and progress monitoring data, and appeared to possess a basic understanding of RTI-related principles. However, responders also indicated more specific guidance and communication that detailed how they could implement the system in their educational setting would have benefitted and perhaps even accelerated the multi-tiered system scale-up efforts.

Shepherd and Salembier (2010) qualitatively investigated the scale-up and implementation of a multi-tiered framework in a northeastern rural elementary education setting and also noted the positive effect the model seemed to have on student reading outcomes. The researchers shared the collaboration and communication that occurred during grade-level universal screening and progress monitoring data dives led to a school-wide pedagogical methods revision and helped to redefine and expand the roles and responsibilities of both general and special educators. Data also revealed the multi-tiered implementation helped establish a school-wide focus on literacy and led to the establishment of both a school leadership team and a professional learning community. As a follow-up to the original study, Shepherd and Salembier (2011) expanded their participant pool to three schools and found multi-tiered scale-up efforts in those settings also led to increases in collaborative data-based decision making; spurred conversations that helped develop a universal, commonly-shared level of understanding; sparked professional changes and responsibilities for principals; and contributed to improved student reading outcomes.

Many of the researchers noted the positive impact of communication, collaboration, and professional support on the multi-tiered instructional initiative by specifically examining teacher perspectives. Swanson, Solis, Ciullo, and McKenna (2012) designed their two-year study to qualitatively investigate the professional experiences and personal perceptions of a small sample (n = 17) of special educators related to a multi-tiered system of supports. Data were gathered during focus group interviews and classroom observations. Results revealed participants felt the multi-tiered model (a) positively impacted their ability to identify student needs and target their instruction, (b) provided increased opportunities to collaborate, and (c) increased both student and teacher levels of engagement. Pyle, Wade-Woolley, and Hutchinson (2011) also noted the importance of communication, collaboration, and professional support during multi-tiered model scale-up efforts. The qualitative study by Pyle et al. investigated the perspectives of 13 educators from five schools in Ontario, Canada related to an initial multi-tiered systems scale-up effort. Thematic analysis and subsequent interpretation supported the notion that when bringing a multi-tiered educational initiative to scale, teachers preferred to work collaboratively and wanted to feel they (a) were involved in the process, (b) had real and meaningful roles, (c) received support from peers and leaders alike, (d) were provided with relevant professional development, and (e) were kept informed on the progress. Similarly, Wilcox, Murakami-Ramalho, and Urick (2013) qualitatively investigated the perspectives of 117 general education teachers in Texas and Michigan on the multi-tiered system of supports framework. The thematic analysis confirmed the importance of (a) providing relevant and timely professional development, (b) collaboration, and (c) having the professional knowledge and skills needed to analyze and use student-level data during the planning process.

To build a better understanding of educators' perspectives related to multi-tiered systems, Scanlon (2013) shared the results of an electronic survey distributed to a large sample of reading teachers and/or literacy coaches (n = 2,700) by the International Reading Association (IRA). The survey asked participants to focus primarily on the scale-up efforts of a multi-tiered system in first grade and share their general perceptions of the multi-tiered model. Similar to findings reported by earlier research (e.g., Swanson et al., 2012; Wilcox et al., 2013), nearly 70% of the responders indicated the multi-tiered system had increased the collaboration in their building and the initiative had a positive impact on literacy instruction.

Some researchers found MTSS scale-up efforts meant both teachers and administrators needed to be prepared to cooperate and assume different roles and responsibilities during scale-up efforts. For example, Bean and Lillenstein (2012) sent a questionnaire to five elementary school principals who worked in schools that had been using a multi-tiered system of supports for at least five years. The authors gathered additional data by conducting classroom observations and interviewing a variety of individuals with diverse roles and responsibilities in each school. The qualitative analysis revealed how the implementation of a multi-tiered model had necessitated a shift in professional roles for a variety of the study's participants. Respondents also noted strong interpersonal and communication skills were required to scale-up a multi-tiered system of support to (a) establish trust, (b) engage in problem solving conversations, (c) collaborate with team members, and (d) provide difficult feedback.

The above research detailed how technical support, effective communication, and collaboration positively impacted the scale-up and implementation efforts of multi-tiered systems of support. For an example of an unsuccessful implementation effort, consider a study conducted by Orosco and Klinger (2010). The authors used qualitative case study methods to examine why a multi-tiered system of support failed during scale-up efforts at a school with a large population of Latino English learners who were struggling while learning to read. The qualitative thematic analysis revealed the scale-up effort suffered from (a) a misalignment between assessment data and instruction, (b) a negative school culture, (c) challenges related to insufficient professional development and educator support, and (d) limited resources that combined to negatively impact student literacy outcomes.

A variety of researchers examined the scale-up of a multi-tiered system using mixed methods. For example, a well-designed quasi-experimental study conducted by Dougherty Stahl, Keane, and Simic (2013) used mixed methods to study the pilot implementation of a multi-tiered system of supports using three first-grade classrooms in an urban school district. Two of the schools received the support and technical assistance of a RTI facilitator while the third did not. Results revealed student risk levels on measures of phonemic awareness and decoding tasks from the dynamic indicator of basic early literacy skills (DIBELS; Good & Kaminski, 2002) decreased in all schools but results favored the schools that received technical assistance and coaching support. Qualitative data gathered by the research team suggested surface-level changes were operationalized in the schools that had received technical assistance in the first year but failed to identify more complex and comprehensive system-level changes. However, Dougherty Stahl et al. did find project participants had increased their (a) skills with assessment, (b) abilities to provide differentiated instruction, (c) expertise using data to guide instruction and make decisions, (d) collaborative competencies, and (d) reflective practice skills. Similarly, research conducted by Harn, Chard, Biancarosa, and Kame'enui (2011) examined the scale-up of a multi-tiered model in two school districts located in the Pacific Northwest using quasi-experimental methods. Specifically, the authors evaluated the effect of providing coordinated, aligned, increasingly intensive, and targeted interventions on student reading outcomes with an uncoordinated effort. Results revealed that when grade-level teams communicated and collaborated to provide students with interventions that aligned with the classroom instruction, those efforts had modest but practically significant effects on student reading outcomes.

Family, School, and Community Partnerships

Family, school, and community partnering (FSCP), according to the Colorado Department of Education (CDE; 2016), can be used to describe what happens when families, school professionals, and community members actively communicate and collaborate to improve learning. For example, formal and informal partnerships between parents and classroom teachers are created at the start of each school year in order to build a positive learning environment that helps propel student learning forward. Additionally, universities, districts, and individual schools frequently partner and collaborate with each other to train and mentor pre-service teachers during their student teaching experiences. A wide variety of examples can be used to describe educational partnerships. Unfortunately, there has not been a great deal of examining how those collaborative partnerships impact student reading outcomes. However, the research that has been conducted can be used to develop a more complete and comprehensive understanding of the role collaborative partnerships play in multi-tiered system scale-up efforts. In the following narrative, study summaries that detail the purposes, participants, methods, and results are provided.

A small group of researchers examined preservice teachers' knowledge of multi-tiered systems, the methods educator preparation programs used to develop that knowledge, and reported mixed results. McCombes-Tolis and Spear-Swerling (2011) studied how thoroughly institutions of higher education with educator preparation programs prepared pre-service teachers to serve elementary students in a multi-tiered system of support model by collecting a sample of 29 reading course syllabi. Using the contents of the syllabi as a guide, the authors concluded preservice teachers were not consistently being prepared to understand key terms, concepts, and pedagogical practices associated with a multi-tiered instructional model. Similarly, Barrio and Combes (2015) examined the concerns of preservice teachers related to a multi-tiered instructional model. The results suggested the preservice teachers who participated in the study felt unprepared to meet students' needs and did not believe they would have the skills or knowledge to effectively implement a multi-tiered instructional model after they graduated.

A sample of research findings provided evidence that collaborative partnerships could be used to improve preservice teachers' MTSS-related skill sets. Hoppey (2013) described how one educator preparation program used action research to (a) help preservice teachers develop a deeper knowledge of the RTI framework, (b) become familiar with the key concepts of a multi-tiered model, and (c) understand how to use student-level data to both make placement decisions and inform lesson planning activities. After the action research project, the participants reported they felt more confident and capable of meeting diverse students' needs. More recently, Mokhtari, Neel, Kaiser, and Hong-Hai (2015) designed a study that provided a sample of first-grade students with a Tier II intervention, offered pre-service teachers with an opportunity to practice their budding pedagogical crafts, and helped develop a university-district partnership. The authors and principal investigators provided the preservice teachers with ongoing and intensive support to develop their skills (a) using evidence-based practices, (b) analyzing data to make instructional decisions, and (c) organizing and planning small-group instruction. At the end of each day, the preservice teachers were also provided with the opportunity to collaborate with their peers and the authors. Results demonstrated positive effects of the partnership and intervention on both the students and teacher candidates. Mokhtari et al. shared the partnership (a) helped preservice teachers and authors gain access to student-level benchmarking and progress monitoring data, (b) facilitated the early identification of students who needed supplemental reading support, (c) created a positive school culture and climate that maximized the impact of the intervention, and (d) was used to facilitate communication efforts between all parties involved (preservice and in-service teachers, students, parents, and faculty).

In addition to school and community partnerships, the role of the family in an effective scale-up cannot be underestimated. In a meta-analysis of the literature that examined the effect family-based reading interventions had on students' reading skills, Senechal and Young (2008) summarized the findings of 16 studies published between 1970 and 2005. The authors noted the studies cumulatively found that high levels of parent-involvement had a positive impact on reading achievement.

In summary, five overarching themes emerged during the research review that examined the impact of providing technical support, facilitating communication, and developing collaborative partnerships on elementary students' reading achievement. First, many of the researchers identified the importance of developing a positive climate at grade-, school-, and district-levels (e.g., Orosco & Klinger, 2010; Scanlon, 2013). Because MTSS is a school-based reform initiative, a variety of school-based professionals must work together to successfully bring the model to scale. Further, the National School Climate Center (NSCC; 2015) shared the way people feel about being in schools has an impact on student learning and development. When groups of people work together in a positive school climate, academic achievement outcomes are positively impacted (NSCC, 2015). Researchers also found MTSS leaders must make sure to include a variety of educational professionals to drive implementation efforts that include classroom teachers, specialists, special education teachers, and parents (e.g., Bean & Lillenstein, 2012; Dougherty Stahl et al., 2013). According to Kezar (2009), collaborative endeavors that include a wide variety of voices and viewpoints tend to maximize student success. Additionally, identifying clear roles, responsibilities, goals, and student-level outcomes in most instances appeared to facilitate implementation efforts (e.g., Shepherd & Salembier, 2010). When teachers and school-based professionals know what the goal is, communicate with each other about the plan, and feel like they can rely on their teammates in real and practical ways, a trust-filled environment will develop, which ultimately leads to improved student learning. Further, the review demonstrated that institutions of higher education need to do a better job preparing preservice teachers for the rigors of everyday classroom practice. Special focus should be paid to teaching preservice teachers to understand the elements of a multi-tiered model and differentiating instruction for students whom they will serve after attaining licensure (e.g., Bario & Combes, 2015, McCombes-Tolis & Spear-Swearling, 2011). Finally, the researchers also noted that using student data to spark implementation efforts could be time consuming but are a critical feature of the MTSS initiative (Orosco & Klinger, 2010; Scanlon, 2013). Data used to identify specific learning targets and track how much student progress has (or has not) been made take some of the guesswork out of improvement efforts. In sum, the findings of the studies noted the positive impact technical assistance, effective communication, and professional collaboration and partnerships have on MTSS scale-up efforts and student learning outcomes.

Universal Screening and Progress Monitoring

As noted previously, the focus of RTI and MTSS has shifted from an alternative special education identification tool to a method that facilitates early identification of students at risk of developing academic difficulties while simultaneously building the capacity of educators and administrators at classroom-, grade-, school-, and district-levels. Student-level data obtained by universal screening and progress monitoring measures are the primary tools educators and administrators alike must use to determine if instructional strategies being used in universal Tier I settings are effective and to identify students who might benefit from an additional level of support (Fuchs & Fuchs, 2006; Jenkins, Hudson, & Johnson, 2007).

Various universal screeners and progress monitoring tools are routinely used by school-based professionals to identify students at risk of developing reading difficulties. One commonly used universal screening and progress monitoring tool is DIBELS (Good & Kaminski, 2002; see Appendix A for more detail). The AIMSweb (NSC Pearson, 2014) and the Gates-MacGintie Reading Tests (GMRT; MacGintie & MacGintie, 2006) are other universal screening and progress monitoring tools commonly used in educational settings to assess students' reading skills and monitor responses to tiered interventions and supports. For an excellent and comprehensive list of universal screeners that includes classification accuracy ratings, the level of generalizability, reliability and validity estimates, and efficiency data for each measure, see the screening tools chart originally developed by the Center for Response to Intervention at the American Institutes of Research included in Appendix B.

The body of research that grounds the utility of using universal screening measures and progress monitoring tools is deep and rigorous. Summaries of quantitative studies that employed experimental and quasi-experimental methods are included. A small sample of qualitative studies is also included to attempt to provide the reader with a comprehensive and complete understanding of universal screening and progress-monitoring data research. This section concludes with a summary of the overarching themes and findings from the reviewed literature.

Universal Screening and Progress Monitoring Protocols

Some of the earliest MTSS researchers investigated assessment protocols that could be used to reliably identify students at risk and provide them with supplemental interventions and instructional supports. For example, the purpose of a study conducted by Al Otaiba and Fuchs (2006) was to identify individual student characteristics that could predict students who would not respond to intensive supports and interventions offered in a supplemental small group setting. Al Otaiba and Fuchs recruited 104 students who were assessed in kindergarten and first grade with a variety of academic and behavioral screeners to participate in the study. Results revealed a combination of letter naming speed, vocabulary, sentence imitation, problem behavior, and the quantity of interventional services each student received predicted 82% of students who failed to respond to supplemental supports, 30% of students who responded occasionally, and 84% of the always-responsive students.

A second early study conducted by Linan-Thompson, Cirino, and Vaughn (2007) explored screening data of 142 first-grade English learners (EL) at risk for developing reading difficulties. Results of this study indicated students' first-grade progress monitoring data were most predictive of students' universal screening benchmarking scores at the beginning of second grade. However, Boscardin, Muthen, Francis, and Baker (2008), concerned with the serious theoretical and technical problems related to identifying students with reading difficulties and/or disabilities, developed a new assessment and screening protocol they claimed could be used to both reliably identify and assess the progress of students at risk of developing reading difficulties or disabilities. Boscardin et al. used existing screening data of 411 primary students and found phonemic awareness (PA) and rapid letter naming were highly predictive of later word recognition and reading skills.

A variety of researchers wanted to discover if benchmarking data gathered using a variety of universal screeners or if progress monitoring data were more reliably predictive of students' future reading scores. Vellutino, Scanlon, Zhang, and Schatschneider (2008) explored if progress monitoring data could be used to differentiate students who continued to be at risk from students no longer at risk when compared to the benchmarking data gathered using universal screeners, measures of intelligence, or measures of reading-related skills. The results indicated progress monitoring measures more effectively and consistently distinguished between these two groups than psychometric measures (e.g., measures of intelligence, or reading-related cognitive abilities). Similarly, Schatschneider, Wagner and Crawford (2008) conducted a large-scale, multi-year study to investigate whether using students' initial academic achievement status, rate of reading growth, or the two sources of data combined predicted future reading achievement using a large sample of first graders (*n* = 23,438). However, the results of this study indicated students' initial achievement status was a better predictor of future reading achievement than reading growth data alone. Findings from these early studies contradicted each other and did little to help the field clarify the types of assessment data that could be used during the instructional planning process in schools striving to bring a multi-tiered system of supports up to scale.

Because previous research findings failed to conclusively identify measures that could be used to predict students' future reading achievement, various researchers continued to examine a range of assessment protocols that might prove to be reliably predictive. One example of this type of study was conducted by Fuchs, Compton, Fuchs, Bryant, and Davis (2008). The researchers used the findings from a series of large-scale, longitudinal, and experimental studies and found measures of word identification fluency (WIF), letter-sound matching, rapid digit naming, oral vocabulary, and WIF progress monitoring scores could also be used to identify students at risk of developing reading disabilities. Using the same data, Fuchs et al. found student scores on measures of sight word reading efficiency, WIF progress monitoring scores, and the discrepancy between oral reading fluency (ORF) rates and WIF progress monitoring scores were equally predictive of students' future reading abilities. In research conducted by Chard et al. (2008), the research team also used an existing longitudinal data set to identify individual reading measures schools could theoretically utilize to predict elementary students' future reading skills using the DIBELS universal screener (Good & Kaminsky, 2002). Chard et al. used the beginning of year benchmark screening data of 668 students from Oregon and Texas and the results of their analysis suggested scores on early screening measures related to alphabetic principles could reliably predict students' ORF rates. Similarly, Hagans (2008) investigated the validity of two DIBELS subtests that measured student skills with phoneme segmenting and basic phonics to determine if they could also be used to accurately predict early literacy skill attainment. Hagans identified a sample of 75 first grade students with a low socio-economic status and found the two DIBELS measures successfully predicted student reading growth. When viewed in total, the results of Fuchs et al., Chard et al., and Hagans seemed to indicate student benchmarking and progress monitoring data gathered using universal screeners had the potential to accurately identify at-risk students. However, in a study by Catts, Petscher, Schatschneider, Bridges, and Mendoza (2009), the researchers wanted to confirm that DIBELS data (Good & Kaminski, 2002) could be used to accurately identify students who were at risk. Catts et al. were curious because previous researchers found DIBELS and similar tools (e.g., AimsWeb) tended to have high levels of over and under identification (e.g., Glover & Albers, 2007; Jenkins et al., 2007). To explain, when a universal screener tends to under-identify students, educators might overlook students at risk of developing reading difficulties and would benefit from an additional, increased level of support. Conversely, when a universal screener tends to over-identify students, educators might incorrectly place students not at risk in an intensive small group for supplemental instruction when those services are not required. Therefore, Catts et al. identified a large sample of 18,667 students who started kindergarten during the 2003/2004 academic year and used their raw benchmarking scores from a range of DIBELS subtests as independent variables and end of year third grade DIBELS ORF scores as the dependent variable. Results demonstrated the universal screening data gathered using DIBELS measures tended to over-identify students and negatively impacted the predictive validity of the measure, which directly contradicted earlier researchers' findings (e.g., Chard et al., 2008; Hagans, 2008).

Attempting to provide support and clarity to the growing confusion surrounding universal screening and progress monitoring data, Deno et al. (2009) described the development of an assessment protocol that was part of an multi-tiered systems framework at the elementary level. In this project, faculty members and graduate students from an institution of higher education collaborated with classroom educators and school-level administrators to design and employ a unique assessment process to (a) analyze within-year reading growth of students, (b) examine across-year contrasts for all students, (c) efficiently administer universal screeners to all students to identify those at risk of developing a reading disability, and (d) provide meaningful opportunities for classroom educators to engage in the process. Study results revealed data from a fall administration of a silent reading comprehension measure reliably identified students at risk and demonstrated that fall benchmarking data could be used by classroom teachers to set future reading goals. The results from Deno et al. seemed once again to contradict the findings of Catts et al. (2009) and further confused the field on the role of universal screening in multi-tiered systems scale-up efforts.

Confusion surrounding screening and assessment protocols continued in later studies. For example, Fuchs, Compton, Fuchs, Bouton, and Caffrey (2011) studied the construct and predictive validity of an assessment protocol called dynamic assessment (DA). A sample of first grade students (*n* = 318) was tested using various individual measures that were combined to create the DA instrument. Fuchs et al. wanted to ascertain if the DA assessment instrument and protocol could be used to accurately predict students' response to increasingly intensive interventions and supports in a multi-tiered system of supports framework. Results revealed the DA instrument reliably predicted student responses to tiered supplemental supports and contributed a unique variance to end of first grade word identification fluency and reading comprehension data above and beyond similar reading benchmarking tools.

Predictive Qualities of Screening Tools

Various additional researchers focused on the predictive qualities of standard universal screening and progress monitoring data. For example, the purpose of a study conducted by Speece et al. (2011) was simply to identify a screening protocol that could accurately and efficiently categorize first grade students at risk of developing reading difficulties using a sample of 243 children. Speece et al. found the best predictors included in the protocol tested students' abilities with word identification. Additionally, in a study conducted by Clemens, Hilt-Panahon, Shapiro, and Yoon (2012), the authors examined screening data from letter naming fluency (LNF), nonsense word fluency (NWF), initial sound fluency (ISF), and phoneme segmentation fluency (PSF) subtests of DIBELS (Good & Kaminski, 2007) to determine if some would be better predictors of first grade end-of-year benchmarking scores than others. Using the universal screening data from a sample of 101 kindergarten and first grade students, the authors found LNF and NWF were more accurate predictors than ISF or PSF, which could be used by teachers to differentiate between students who struggled and those who would be able to meet end-of-year reading expectations.

Similarly, Compton et al. (2012) wanted to identify an assessment protocol that could accurately identify children who would be unresponsive to supplemental, tiered supports and should move directly from Tier I to Tier III or special education. To investigate, Compton and his colleagues identified a total sample of 129 first grade students who had been unresponsive to classroom-level Tier I instruction and used random assignment to select a subgroup of students who received 14 weeks of supplemental Tier II supports. Compton et al. then used end-of-year second-grade benchmark criteria to identify the students who were either responsive or non-responsive to the Tier II intervention. Results revealed Tier II response data were not needed to correctly classify students in terms of response and nonresponse status. Once again, these findings contradicted earlier studies and confounded the educators and administrators as they tried to identify students at risk.

Unfortunately, this lack of clarity on the utility of using universal screeners to predict reading achievement outcomes has continued. Various researchers have investigated the impact of universal screening data that falsely identified students as "at risk." For example, in a study conducted by McAlenney and Coyne (2015), the authors attempted to develop a tool that could minimize the amount of students incorrectly categorized as at risk of developing reading difficulties early in kindergarten. The researchers assessed a sample of 105 kindergarten students with beginning-of-year screening data that suggested they were at risk and, as a result, were placed in a year-long Tier II reading intervention. Students (n = 9) with very robust curriculum mastery scores were identified as possible false positives and were removed from the Tier II intervention group. During the end-of-year benchmarking window, this group of students scored above the at-risk level and functioned similarly to the group who remained in the Tier II group for the duration of the academic year. Overall, while a considerable body of universal screening and progress monitoring research has accumulated since the reauthorization of IDEA in 2004, little clarity has been gained by schools about the best way to efficiently and accurately predict and identify students who would benefit from supplemental academic interventions of a multi-tiered system of supports.

Uses of Screening Data

Some researchers have tried to understand how universal screening and progress monitoring data are being used by teachers. For example, Jacobs, Gregory, Hoppey, and Yendol-Hoppey (2009) sought to understand the unique ways classroom educators use data to inform their instructional planning with a sample of nine elementary teachers. Qualitative methods were used to identify nine individual themes. First, the participants shared that using data in meaningful ways required a high degree of professional knowledge and expertise. The participants also shared that using data helped them as they focused on the needs of individual students, created a sense of instructional urgency, and propelled the decision-making process forward. Finally, the participants shared that student data were generally used during the instructional planning process, served to advance their professional knowledge, and engendered a culture of support and collaboration.

When viewed comprehensively, the above studies united under three themes. The largest proportion of the researchers wanted to discover if universal screening and/or progress monitoring data could be reliably used to identify students at risk and make predictions of future reading scores (e.g., Al Otaiba & Fuchs, 2006; Linan-Thompson et al., 2007). Historically, the body of research demonstrated the longer students struggled while learning to read, the more difficult it became for them to catch up and keep up with their grade level peers. Therefore, to maximize both the reading outcomes of young students as well as the long-term impact on student learning of the multi-tiered system, educators must be able to access valid and reliable student-level data to both identify students at risk and intervene early (Wanzek & Vaughn, 2007). A second and related theme centered on the identification of individual reading measures that could be included in an efficient assessment protocol that would generate reliably predictive data (e.g., Chard et al., 2008; Fuchs et al., 2008). In a multi-tiered model, data help to drive the instructional planning process, evaluate the effectiveness of a program, and inform school-improvement efforts. One could argue that in a multi-tiered model, data fuel the engine of learning and improvement and few contend that using data is optional. However, it should be obvious that data must be gathered in in an efficient and efficacious manner because teachers rarely have large blocks of unscheduled free time built into their daily instructional schedules. Recognizing this fact, the researchers attempted to clarify the screening and/or progress monitoring measures that could be included as part of an efficient and effective assessment protocol. Unfortunately, instead of clarifying, the work conducted by the researchers only confused and confounded what the field understood about efficiency, validity, and reliability. Future research might provide more helpful information. Finally, perhaps the most important idea identified during the review of the research focused on the assessment literacy knowledge and skills of teachers. Defined by Popham (2009) as the level of understanding and expertise teachers have with the basic concepts of classroom-related measurement, in todays' era of accountability-driven education, assessment literacy-related skills are a mandatory requirement rather than an option. Therefore, various researchers were interested in discovering how teachers gathered and used universal screening and progress monitoring data and how those practices impacted the scale-up efforts of a multi-tiered model (e.g., Deno et al., 2009; Jacobs et al., 2009). Cumulatively, the results indicated that while teachers understood the importance of data-driven instruction, they also lacked the assessment literacy-related skills and professional confidence needed to independently gather, interpret, and use student-level data during instructional planning. As noted in the previous section, scaling-up a multi-tiered system effectively means teachers must receive timely and relevant technical assistance. Research included in this portion of the synthesis indicated future research should focus on developing assessment literacy skills of classroom educators so they can confidently gather and use student-level data.

Evidence-Based Practices

The Colorado Department of Education (2016) defined evidence-based practices (EBP) as "approaches to instruction, intervention, and assessment that have been proven effective through research indicating improved outcomes for students" (p. 1). A wide variety of researchers who investigated the methods and concepts classroom educators should use to stimulate students' reading growth over the past 35 years developed a strong consensus about specific components that served as the foundation of effective early reading pedagogy (NRP, 2000). As shared earlier, reading instruction that builds student skills and knowledge with phonemic awareness, phonics, fluency, and vocabulary, and develops students' skills with a range of comprehension strategies are more effective than those that do not (Snow, Burns, & Griffin, 1998). Unfortunately, an examination of that significant and deep body of research did not align with the topic and scope of this review. However, to ignore evidence-based practices completely while discussing reading achievement is neither warranted nor wise because the MTSS framework was developed to positively impact students' academic and behavioral outcomes by building systems that support and develop teachers' knowledge. Therefore, this segment briefly summarizes the body of research that specifically examined educators' knowledge and pedagogical skills related to reading. Similar to previous sections, this review provides a brief summary of each study and concludes by providing a summary of themes that united these investigations.

Researchers have consistently found that providing students with differentiated instruction in universal or Tier I settings positively impacts student reading outcomes. For example, in a study conducted by Menzies, Mahdavi, and Lewis (2008) that evaluated the reading progress a sample of first grade students made after receiving a tiered intervention, the research team incorporated a range of instructional strategies targeted to meet student needs in their design and provided teachers with ongoing technical assistance and on-site coaching support. The instructional coaches supported teachers while they learned to (a) gather progress monitoring data to assess student growth and skill acquisition, (b) provide high intensity instruction to students who were at risk in supplemental small groups, (c) use explicit instructional strategies with students lacking in PA skills or who did not seem to grasp the basics of the alphabetic principle, and (d) collaborate with one another and a literacy coach. Results indicated that helping teachers gain the above skills positively impacted students' reading achievement as 90% of the students met or surpassed grade-level benchmark expectations by the end of the academic year.

Rodriguez and Denti (2011) took a more general approach and studied how using a prepared curriculum, gathering student progress monitoring data, and using the data to guide instructional planning activities impacted the reading outcomes of second grade English learners. Results of the study indicated students whose teachers (a) used a commercially-prepared evidence-based curriculum but (b) knew how to monitor their students for progress, and (c) used data to make instructional planning decisions had significantly higher rates of growth on measures of ORF than the students whose teachers did not.

Other researchers found the amount of MTSS-related knowledge teachers possessed had a direct impact on reading-related multi-tiered systems scale-up efforts. For example, Spear-Swerling and Cheesman (2012) examined teachers' basic skills and level of knowledge as they implemented multi-tiered models of instruction in reading. A multiple-choice survey was developed by the research team to measure participants' knowledge of (a) the individual reading components identified by the NRP (2000), (b) methods of assessment, and (c) generalized multi-tiered practices. The researchers distributed the survey to K-5 elementary teachers within the sampling frame and received responses from 142 individuals. Results revealed participants were most familiar with fluency, vocabulary, and comprehension; they were least familiar with assessment protocols and RTI-related practices. While the teachers seemed more familiar with the evidence-based elements of reading instruction, their ability to incorporate that information into a multi-tiered model seemed limited. Similarly, Jenkins, Schiller, Blackorby, Thayer, and Tilly (2013) gathered data that allowed them to report how the scale-up efforts of multi-tiered systems in school-based settings differed from the research-based recommendations made by Gersten et al. (2009). Jenkins et al. distributed a survey to a sample of 62 elementary teachers, which permitted the researchers to evaluate (a) the extent student data were used to identify students' reading risk levels and guide placement decisions and (b) the ways schools with more experience using a multi-tiered framework differed from schools who were just beginning their scale-up efforts.

While results unsurprisingly indicated scale-up efforts varied widely, most teachers indicated they used a multi-tiered framework to provide students with increasingly intensive supports in at least two subjects and used curriculum-based measures to gather student-level data. Participants also indicated supplemental tiered supports were provided to students an average of four to five days per week and small teacher to student ratios were used when building student intervention group rosters. However, schools varied widely in the amount of time students spent per week receiving supplemental supports and interventions. Finally, while the authors anticipated schools with more multi-tiered systems experience would have developed complex models with intricate program architectures to provide students with tiered supports across a range of contents and grade-levels, results failed to identify any differences. The findings from these two studies seemed to indicate that while classroom teachers who were working in real-world settings possessed a basic understanding of the essential elements of both reading instruction and of a multi-tiered model, a deeper conceptual grasp of the potential of the multi-tiered model had not yet surfaced.

More recently, researchers have started to find teachers' knowledge of and skills with multi-tiered systems are improving. For example, Regan et al. (2015) examined the perceptions of 63 teachers in a northeastern school district that had recently adopted a multi-tiered system of supports framework. Using an electronic survey, participants were provided with the opportunity to share their opinions on (a) the feasibility and effectiveness of using evidence-based practices and progress monitoring data, (b) their knowledge of basic RTI concepts, and (c) their perceived ability to implement individual components of the RTI framework into their instructional practices. Results of this study revealed participants felt well prepared to use evidence-based instructional practices and progress monitoring data. Additionally, most participants indicated they understood the basic principles of RTI but also felt they would benefit from direction and guidance during the implementation process. In one of the most recent studies conducted by Al Otaiba et al. (2016), the research team examined the changes in kindergarten teachers' ability to provide differentiated Tier I instruction as a result of a two-year professional development opportunity and analyzed how that training impacted student reading and vocabulary development. The researchers recruited a sample of 10 teachers who served 416 kindergarten students in four separate schools. Findings indicated the teachers provided more differentiated instruction and students had higher word reading outcomes after the professional development program than they did prior to the training.

The themes associated with this portion of the literature review either added to or built on those identified in earlier sections. First, a variety of researchers continued to confirm the importance of providing teachers with the opportunity to collaborate and receive technical support (e.g., Menzies et al., 2008; Rodriguez & Denti, 2011). As noted earlier, successful scale-up efforts of multi-tiered models that positively impact student reading outcomes must provide opportunities for teachers to work together and ensure they receive timely and appropriate technical assistance and support. Secondly, it appeared teachers' knowledge and skills related to evidence-based reading practices and the basic concepts associated with multi-tiered models tended to improve over time (Jenkins et al., 2013; Regan et al., 2015; Spear-Swerling & Cheesman, 2012). In the same way teachers provide their students with time and instructional support to learn a concept before mastery is the expectation, educational policy makers and federal legislators must also recognize that classroom teachers need both coaching support and time to master the concepts associated with multi-tiered systems. As research has indicated, while teachers seemed to indicate a strong desire to learn how to incorporate the elements of a multi-tiered model into their educational environment, they needed time and support as they learned how to do so (e.g., Al Otaiba et al, 2016; Deno, 2009).

Layered Continuum of Supports

According to Gelzheiser, Scanlon, Vellutino, Hallgren-Flynn, and Schatschneider (2011), to bring a multi-tiered system up-to-scale and provide supplemental supports capable of accelerating the learning of students at risk, schools should be using evidence-based practices provided within a layered continuum of supports. Oakes et al. (2014) refined this definition by sharing that a layered continuum of supports and prevention has the potential to improve long-term academic, social, and behavioral outcomes for a full range of students including those with and without disabilities. Currently, the Colorado Department of Education (2016) defines the layered continuum of supports as a system designed to ensure

> that every student receives equitable academic and behavioral support that is culturally responsive, matched to need, and developmentally appropriate, through layers that increase in intensity from universal (every student) to targeted (some students) to intensive (few students). (p. 1)

Since the inception of multi-tiered systems like RTI, researchers have investigated how providing students with increasingly intensive interventions offered in either the classroom environment to every student or to specific groups of students in supplemental Tier II and Tier III settings impacted students' reading achievement. Because relevant studies that examined Tier I instruction have been previously described and summarized, this section primarily focuses on studies that investigated the impact of Tier II and Tier

III interventions on student reading outcomes. Similar to earlier sections, this review shares details of experimental, quasi-experimental, and qualitative studies that have informed the field. This section once again ends with a thematic review and analysis.

Findings from some of the earliest research investigated the effects of providing elementary students with supplemental supports as they developed their skills with reading. A meta-analysis of the research literature completed by Wanzek and Vaughn (2007) examined the impact of providing early reading interventions to students who were at risk and shared the findings of 18 studies that met inclusion criteria of being published between 1995 and 2005. The combined findings of the studies supported ideas that intervening with students who were at risk (a) in groups with small student-to-teacher ratios, (b) during earlier grades (K-1), and (c) with fluctuating levels of intensity that varied by time and/or duration had the potential to positively impact student reading achievement.

Using this earlier research as a foundation, most of the recent experimental and quasi-experimental studies conducted by researchers compared the reading achievement of students who were provided with supplemental supports in tiered settings with the reading achievement of students who remained in Tier I. Cumulatively, the researchers found providing students who were at risk with supplemental supports in small-group settings had a positive impact on reading outcomes. For example, Schuele et al. (2008) conducted a study that was focused on developing kindergarten students' PA skills with a sample of 113 students. Students in the control group were provided with Tier I instruction while students in the treatment group were provided with Tier I instruction supplemented with either a classroom-based PA instruction or a 12-week Tier II small group intervention. Results indicated that providing the supplemental PA in the Tier I classroom setting did not produce statistically significant gains for typically achieving students on measures of letter sound knowledge, word recognition skills, or spelling. However, when the Tier II small group intervention was added, students who were at risk outperformed their control group peers.

Similarly, Case et al. (2010) used experimental methods to validate a short-term Tier II reading intervention with a sample of first-grade students at risk of developing reading difficulties (n = 30). Students in the treatment group were provided with a 16-hour intervention in a small-group setting that focused on developing students' PA, word attack skills, spelling, sight word recognition, vocabulary knowledge, oral reading fluency, and comprehension abilities. Students in the control group remained in Tier I. Results suggested the short-term supplemental reading interventions had significant, positive effects on the reading skills of students in the treatment group when compared to students in the control group. Most recently, Baker, Smolkowski, Chaparro and Fien (2015) examined the effect of providing first-grade students at risk with both Tier I and Tier II reading supports and Tier I-only instruction using regression discontinuity methods. Results suggested students at risk who received Tier I and Tier II supports made greater reading gains than students who only received Tier I supports.

While the body of research consistently supported the effectiveness of providing students at risk with supplemental supports in small group settings, the findings were mixed on the amount of time students at risk should be provided with Tier II interventions. For example, in an exceptional randomized control trial, Vaughn et al. (2009) investigated the effects of providing an intensive Tier II reading intervention to a sample of students who demonstrated minimal response to previous, less intensive supports. After identifying an initial sample of 274 first-grade students who met the researchers' criteria for first-grade reading difficulties, Vaughn et al. randomly assigned students to either a treatment or comparison group. First grade students in the treatment group received Tier II interventions between 13 and 26 weeks. Students in the control group remained in Tier I. The reading skills of the students in the treatment group were evaluated again at the beginning of second grade. Students meeting expected grade-level benchmarking criteria (higher responders) did not receive further Tier II supports while treatment students who did not meet second grade benchmark reading criteria (lower responders) were provided with an additional 26 weeks of a more intensive intervention. The findings of Vaughn et al. revealed statistically significant differences between students in the treatment condition who received additional supports during their second grade year and those who did not. However, no significant results were identified on measures of oral reading fluency for the lower responder group, which suggested students who were lower responders might need more intensive and long-term interventions.

Similarly, Denton et al. (2011) compared the effects of providing a variety of supplemental Tier II interventions to a sample of first grade students at risk by randomly assigning students to Tier II groups that varied both by frequency and duration. Participants in the extended group (n = 66) were provided with Tier II supports four times per week over 16 weeks. Participants in the concentrated group (n = 64) were also provided with Tier II supports four times per week but the duration of the Tier II group decreased from 16 weeks to eight weeks. In the third group (n = 62), students were provided with Tier II supports two times each week for a total of 16 weeks. Interestingly, the results of this study indicated the groups did not significantly differ on any reading outcome nor identify statistically significant differences in the number of students who failed to respond to the Tier II intervention.

Various researchers examined the impact of providing supplemental supports that varied by intensity, instructional strategies, or focused on one of the five reading components identified by the NRP (2000). For example, Kerins, Trotter and Schoenbrodt (2010) identified a sample of 23 first grade students who were at risk for developing reading difficulties and randomly assigned them to either an experimental or treatment group. Students in the experimental group received 16 hours of Tier II supports that incorporated explicit phonemic awareness training and multi-sensory reading instruction in addition to Tier I instruction. The control group remained in Tier I and received a classroom-based intervention. When comparing the effect of the classroom-based intervention to the effect of the classroom intervention plus 16 hours of additional intensive instruction, the researchers failed to identify any differences between the two groups, which might not be surprising given the small sample size. However, in a randomized control trial, Buckingham, Wheldall, and Beaman (2012) investigated the effectiveness of a Tier II small group reading intervention with a sample of 22 kindergarten students who were randomly assigned to either a treatment or a control group. Students in the treatment group received a Tier II intervention in a small group setting four hours per week for 27 weeks while students in the control group continued in Tier I. Results of this study revealed a large and statistically significant difference between the two groups on post-test measures that supported the efficacy of providing at-risk students with supplemental interventions in small group settings.

Interested in the impact of group size on reading outcomes, Gilbert et al. (2013) identified a sample of students who had been unresponsive to Tier I instruction (*n* = 212) and randomly assigned them to treatment and control groups. Students in the treatment group (*n* = 134) received supplemental Tier II supports in small group settings while students in the control group (*n* = 78) continued in Tier I. Progress monitoring data identified students who were failing to respond to the Tier II intervention (*n* = 45), who were then randomly assigned to receive either more Tier II supports (*n* = 21) or Tier III supports identical to those who were provided in Tier II, but delivered in one-on-one settings (*n* = 24). Results revealed Tier I non-responders who received supplemental supports in Tier II and Tier III settings made significantly higher gains in word reading skills than students in the control group who only received Tier I instruction. However, results did not reveal any differences between students who did not respond to Tier II and subsequently received either more Tier II or Tier III supports. These results suggested that providing Tier III supports identical to those provided in Tier II but only varied by group size might not be as effective as providing students with Tier III intensive supports designed to meet their individual needs.

While most researchers investigated the impact of early interventions with young students in kindergarten and first grade, a variety of others examined how older students responded to supplemental interventions and noted the positive impact those efforts had on students' reading achievement. For example, Ritchey, Silverman, Montanaro, Speece, and Schatschneider (2012) investigated the effect of a Tier II fluency-focused intervention using a sample of 123 fourth grade students who had a high probability of developing reading failure using a randomized control trial methodology. Results indicated fourth grade students who received the Tier II intervention performed significantly higher than their control group peers who remained in Tier I on measures of science and comprehension strategy knowledge and use but not on word reading, fluency, or on general measures of reading comprehension. Additionally, the authors found children with a higher risk level who were part of the Tier II treatment condition appeared to benefit from the Tier II support more than students who were either (a) at lower risk in the treatment group or (b) at higher risk in the control group. Similarly, Graves, Duesbery, Pyle, Brandon, and McIntosh (2011) conducted two experimental studies with a sample of sixth grade students who qualified for free or reduced lunch (n = 109) to evaluate the impact of a Tier II intervention on the sixth grade students' reading skills compared to sixth grade students who remained in Tier I-only settings. In both studies, the Tier II interventions focused on developing students' skills with word analysis, fluency, comprehension, and vocabulary and were provided three hours each week for 10 weeks. Results from both studies indicated significant differences in the fluency scores of the Tier I-only students and the students who received Tier II supports, which favored students who received the supplemental intervention. Additionally, results from the second study indicated students in the multi-tiered condition had significantly higher scores on measures of text comprehension than students in the Tier I-only group.

Other researchers shared the positive impact of providing students from diverse backgrounds with supplemental supports in tiered settings. For example, Linan-Thompson, Vaughn, Prater, and Cirino (2006) compared the reading achievement data from a sample of first grade EL students who were part of a Tier II research-based intervention with the data of EL students who remained in Tier I. Findings revealed more

EL students who were part of the Tier II research-based intervention in first grade were able to meet grade-level expectations by the end of second grade than those who remained in Tier I. Similarly, Calhoon, Al Otaiba, Cihak, King, and Avalos (2007) used an experimental design to examine the impact of a supplemental Tier II peer-mediated reading intervention on the reading achievement of 76 first grade students enrolled in a two-way bilingual education program. Calhoon et al. randomly assigned whole classrooms to either experimental or control groups; students in the experimental classrooms received the peer-mediated early literacy intervention three days a week over the course of 10 weeks. Results of this study provided additional data that supported the positive impact of supplemental interventions.

Other researchers found providing students with challenging behaviors and reading difficulties with supplemental supports in a Tier II setting positively impacted students' reading achievement. For example, Oakes, Mathur, and Lynne (2010) noted students at risk for developing emotional and behavioral disorders historically tended to have low rates of intervention responsiveness. Therefore, Oakes et al. examined the impact of implementing a multi-dimensional Tier II intervention with a group of students who both struggled while learning to read and frequently displayed problematic behaviors. Using a small sample of nine second grade students with challenging behaviors and/or reading difficulties, the results revealed all students benefitted from the Tier II intervention but improved by varying degrees. Oakes et al. noted the students' levels of attention might have contributed to the differences because students with lower levels of attention were less responsive than students with higher levels of attention.

To investigate the responsiveness of students with language difficulties to early reading interventions, O'Connor, Bocian, Beebe-Frankenberger, and Linklater (2010) randomly assigned 78 kindergarten students with poor language skills to Tier II small groups that varied by start dates (beginning-of-year, mid-year). Results demonstrated all students with low language levels or at risk of developing a reading disability benefitted equally from the Tier II intervention irrespective of the time of year the intervention was started.

Similarly, Denton, Fletcher, and Anthony (2006) investigated the effects of a Tier III intervention using a sample of 27 students with severe reading difficulties and disabilities. Participants were provided with a 16-week Tier III intervention that focused on developing student decoding and oral reading fluency skills. The decoding intervention was provided to students two hours each day for eight weeks while the fluency intervention was provided for one hour each day for a total of eight weeks. Results demonstrated that on average, all students realized significant advances in reading decoding, fluency, and comprehension measures.

Denton et al. (2013) conducted a randomized control trial to estimate the effects of an intensive, individualized Tier III intervention for a sample of second grade students (N = 72) who had previously failed to respond to Tier I and Tier II instruction and supports. Students were randomly assigned to either the experimental (n = 47) or to the control (n = 25) groups. Students in the experimental group received an intervention focused on decoding, word recognition, vocabulary, fluency, and comprehension while students in the control group received typical classroom instruction in a Tier I setting. Results revealed students in the experimental group significantly outperformed those in the control group on measures of word identification, phonemic decoding, word reading fluency, text reading fluency, and comprehension at both the sentence and text levels.

The positive impact of providing supplemental and increasingly intensive instruction on the reading outcomes of students at risk was the single overarching theme that united the research summarized in this section (e.g., Schuele et al., 2008; Vaughn et al., 2009). Research that specifically examined the impact of group size, intervention duration, and frequency failed to identify statistically significant differences between groups (e.g., Denton et al., 2011; Kerins et al., 2010). However, when viewed in total, the evidence indicated when students who struggled while learning to read were provided with more opportunities to read in intensive, small-group settings, their reading outcomes improved (e.g., Case et al., 2010; Smolkowski & Cummings, 2015). While most reading experts and educational researchers continue to endorse the idea that providing supplemental supports to younger children during their primary years is the best thing educators can do to stem future difficulties (e.g., Baker et al., 2015; Buckingham et al., 2012), additional results from a sample of studies found older students also responded to supplemental supports (e.g., Graves et al., 2011; Ritchey et al., 2012). Therefore, it can be confidently stated all students who struggle while learning to read could benefit from supplemental tiered interventions provided in small group settings. While those providing those interventions have both financial and instructional implications, the body of research demonstrated that students at risk of developing reading difficulties or who have been identified with a disability could benefit from supplemental supports in increasingly intensive settings.

Analysis of Research Methodologies

As this synthesis demonstrated, a wealth of research investigated the effect multi-tiered systems of support had on elementary students' reading achievement. The studies included in this review used experimental, quasi-experimental, and mixed methods. According to Odom et al. (2005), special education researchers should strive to use a variety of methods because it enables readers to develop a comprehensive understanding of a complex field. Further, high quality research could be used to (a) develop an appreciation of the school, classroom, and societal factors that have an impact on how evidence-based practices function in real-world settings; (b) develop a deeper understanding of the learning environment; and (c) recognize how students use the skills they have acquired. To meet these goals, the Council of Exceptional Children gathered a group of research experts who formed the Task Force on Quality Indicators for Research in Special Education (Odom et al., 2005). Subcommittees of the Task Force developed a set of quality indicators for research methodologies commonly used in special education research: (a) experimental and quasi-experimental group designs, (b) research related to evidence-based practices, (c) single-subject designs, (d) correlations designs, and (e) qualitative research. A discussion that shares how the experimental, quasi-experimental, and qualitative studies measured up to the Task Forces' quality indicators is presented as follows.

According to Gersten et al. (2005), experimental and quasi-experimental methods permit educational researchers to establish whether the implementation of a practice (or educational framework) results in a systematic shift in the outcomes for a specific group of students. To ascertain whether the practice makes a difference, it is used with a group of students and post-test data are compared to the individuals who were part of the control group. Experimental methods are used when participants are randomly assigned to either the treatment or the control groups. Conversely, quasi-experimental designs are employed when it is not possible to use random assignment but the research team can establish that the treatment and control groups have the same set of skills and demographic characteristics. High quality experimental and quasi-experimental studies should address five main categories. First, the researchers must have developed a strong rationale for the study, provided an argument that supported the intervention and composition of the comparison group, and shared the research questions and a clear purpose statement. Secondly, the authors should have adequately described the participants, attrition rates, and characteristics of the individuals who provided the intervention. The authors should have also shared details about the intervention and used a variety of different outcome measures to quantify student responses. Finally, the data analysis discussion should disclose the statistical methods employed to analyze the data, how those methods related to the study's purpose and research questions, and disclose the techniques employed by the researchers to account for the variability within the sample.

For the studies that incorporated a group experimental and quasi-experimental design, in each instance the researcher(s) made a persuasive and convincing argument that the study would help inform the field in a novel way. In many of the studies, the authors used random assignment or close approximations of randomized assignment. While fidelity of implementation information was missing from a large proportion of the reports, the basic elements of the instructional services provided to the participants were thoroughly described. Additionally, most of the authors included in this review made sure to clearly describe the intervention in the body of their reports, provided attrition rates when appropriate and necessary, and detailed the outcome measures used to measure the effect of the intervention. However, in many instances, the authors failed to detail the nature of instruction provided to students in the control group. In fact, many simply described what occurred in the control group simply as “typical school practice.” However, while there is room for improvement, the studies included in this review largely met the quality standards identified by Odom et al. (2005) and described by Gersten et al. (2005).

According to Brantlinger, Jimenez, Klinger, Pugach, and Richardson (2005), qualitative studies use a systematic approach of inquiry to understand the qualities or nature of a phenomenon in specific settings. Odom et al. (2005) shared three principal techniques used in qualitative studies: (a) interviews, (b) observations, and (c) analysis of documents. The quality indicators for interviews state researchers should identify and recruit individuals from the population of interest, create clearly worded interview questions, record and transcribe the interviews, describe the participants in a fair and sensitive manner, and ensure participant confidentiality is maintained. Researchers who engage in observational studies should also identify and recruit individuals from the population of interest, spend enough time in the setting to develop a comprehensive understanding of the environment, strive to be a non-intrusive observer who has a minimal impact on the environment, take comprehensive field notes, and maintain participant confidentiality. Further, Odom et al. noted qualitative research could be evidence-based when the reader (a) could confirm the data collection and analysis methods were satisfactory and sufficient, (b) knew the research practice resulted in significant and valuable changes and/or outcomes for the study's participants, and (c) could ascertain the practice would be useful during future work with a specific population of interest.

Like the experimental and quasi-experimental studies included in this synthesis, the qualitative studies met some, but not all, of the quality indicators identified by Brantlinger et al. (2005) and summarized by Odom et al. (2005). In many instances, the authors did not share the interview questions or disclose how they recorded and transcribed the interviews or focus group conversation. However, in each study, sufficient measures were employed to recruit individuals from the population of interest and maintain participant confidentiality. Additionally, all of the qualitative studies provided relevant and applicable multi-tiered, systems-related data; provided enough detail to affirm the scale-up efforts resulted in significant changes and outcomes; and confirmed multi-tiered system scale-up efforts had the potential to positively impact student reading outcomes. Therefore, all of the qualitative studies included in this synthesis met evidence-based practice quality criteria identified by Odom et al.

Synthesis and Implications for Future Research

The MTSS framework is a recent innovation of the 21st century educational reform movement that combines increasingly intensive student-level interventions of the RTI model with the school-wide focus of the SW-PBIS model. Experts hope the combination of the two models into a single, cohesive framework leads to increased learning outcomes for all students regardless of ability level (e.g., Sugai & Horner, 2009). However, those same experts have not been able to definitively identify individual components of the MTSS framework that lead to increased reading outcomes for students. As demonstrated in this synthesis, a growing body of research has demonstrated the impact universal screening and progress monitoring data have on student reading achievement. Additionally, a wealth of research has demonstrated how providing students with increasingly intensive instruction that is differentiated to meet individual student needs improved the reading outcomes for students at risk. However, researchers demonstrated that preservice and in-service teachers alike seemed to lack the assessment literacy skills it took to efficiently gather and effectively use data to make decisions, solve problems, or revise their pedagogical techniques. These are issues that desperately require more work and research. Additionally, there is a gap in the body of MTSS and reading research investigating how collaborative partnerships that include parents and families impact the reading outcomes of elementary students. Additionally, more research is needed to identify individual components of the MTSS framework that have a meaningful impact on the learning, growth, and development of students and teachers alike. More research is also required to examine how the combination of individual components of a MTSS framework impacts student reading outcomes. It could be argued that the MTSS framework and student reading outcomes are too distal and are impacted by too many variables for any researcher to think about examining how the first affects the second. While that might be true, the field of educational research must also move beyond researching teachers' perspectives on individual elements of the MTSS initiative and instead get to the heart of the matter and discover how the comprehensive framework impacts student reading.

When viewed as a whole, the themes that united the research studies included in this review could be grouped into one of three categories. Throughout the body of MTSS and reading-related research, collaboration was one of the essential elements of successful implementation efforts. Collaborative activities could happen within the school building led by administrators and teacher leaders as they identify common goals and use common language. Collaborative relationships could also happen with families and community leaders as evidenced by the FSCP research. At the most elemental level, collaboration helps facilitate communication, collegiality, and the cohesiveness of implementation efforts.

The second overarching theme that united this body of research was the role student-level data played in any MTSS scale-up effort. In a multi-tiered systems model, data are the fuel that drives implementation and improvement efforts. Because data are a vital component of the framework, it is critically important they are valid, reliable, and gathered in an efficient manner. Researchers consistently demonstrated data could be used to make predictions about student responsiveness and could be used by teachers and administrators to evaluate the effectiveness of an intervention. However, the research also demonstrated that teachers need to learn how to both efficiently and effectively gather valid and reliable data and confidently use data during their instructional planning processes.

The third theme that united the literature included in this review revolved around differentiated instruction. In the MTSS framework, evidence-based practices are used in Tier I environments, which is appropriate for a majority of students. For students who fail to respond to that universal instruction, the MTSS framework could be used to provide additional, more intensive, differentiated instruction to accelerate student learning in supplemental small groups. In summary, collaborative abilities, knowing how to use data, and using differentiated instructional strategies are all critical requirements that would either help or hinder the scale-up of the MTSS educational framework.

CHAPTER III

RESEARCH METHODOLOGY

The previous chapters of this study identified the rationale, purpose, and research questions; and reviewed relevant research that formed the foundation of this dissertation. This chapter begins with a description of the study's design before restating the research questions. A detailed description of the multi-tiered system of supports implementation perception survey (MTSS-IPS; Pierce, Klopfenstein, & Mathis, n.d.), which was used to gather the data for the independent variable, is included. Similarly, a rationale and detailed description of the data source used as the dependent variable--the Transitional Colorado Academic Program (TCAP; CDE, 2014)--are also included. Additionally, the process used to identify and recruit survey participants is described. An overview of structural equation modeling, which was the statistical model used to analyze the data, is included to acquaint the reader with the basic concepts of the statistical model. The final section of this chapter describes the methods employed to handle missing data.

Study Design

A correlational research approach was selected as the most effective statistical method for this study. According to Creswell (2014), correlational research methods have historically used quantitative data and statistical methods to define and assess the degree or correlation between two or more variables or sets of scores. More recent developments in causal designs provide researchers with the opportunity to analyze the relationships among and between a set of variables including hierarchical linear modeling, logistic regression, and structural equation modeling (SEM). The most recent evolution of SEM integrates causal paths and can be used by researchers to collectively detect the strength of a set of multiple variables. A more detailed explanation of SEM is included in later sections.

Research Questions

The following research questions guided this study:

Q1 Does the hypothesized higher order MTSS theoretical factor structure of each measurement model fit the data?

Q1a For the proposed MTSS models hypothesizing relationships between implementation perception of MTSS and 2014 third grade Transitional Colorado Assessment Program (TCAP) data, do the data fit the models?

Q1b Does one model fit the data better than the others?

Q1c What effect does school-level percent of free and reduced lunch have on 2014 third grade TCAP reading scores?

Q1d What effect does district size have on TCAP scores?

Q1e Which latent factors account for more of the variance in student reading outcomes?

Multi-Tiered System of Supports Implementation Perception Survey

According to Fowler (2008), survey research provides researchers with a numerical description of trends, attitudes, perceptions, or opinions of a sample and then uses the results with the aim of generalizing the results to the larger population. Creswell (2014) expanded on this, noting surveys are tools that quantitatively measure “trends, attitudes, or opinions of a population by studying a small sample of that population and generalizing results to the larger group” (p. 155). Groves et al. (2009) also noted surveys systematically collect information from a group to build quantitative descriptions or statistics of the larger population. The survey being used in this study was recently developed by a team of researchers and was designed to measure the six key constructs of the Colorado MTSS model. One of the goals of designing and conducting surveys is to reduce the amount of measurement error those instruments contain (Groves et al., 2009). Measurement error occurs when there are differences between the true answer to an item and the respondent's answer. These differences occur when survey items are poorly constructed, difficult to understand, or easily misinterpreted (Groves et al., 2009). Tourangeau, Rips, and Rasinski (2000) identified five types of comprehension difficulties participants can experience when completing surveys; items can (a) be ambiguous, (b) be too complex, (c) be vague, (d) contain unfamiliar terms, or (e) lead respondents to make false inferences.

Survey Design

The focus of this project was to gather quantitative data enabling participants to provide their perceptions of MTSS implementation efforts in their individual settings. Survey data were gathered using the MTSS-IPS, which was collaboratively developed by a team of researchers at the University of Northern Colorado (UNC) led by Dr. Corey Pierce in partnership with (a) the Office of Learning Supports within the Colorado Department of Education, the (b) Educational Innovation Institute, and (c) UNC's Social Research Laboratory; additional input was provided from a variety of national RTI experts affiliated with the National Center on Response to Intervention (Pierce et al., n.d.). When the MTSS-IPS was being developed, Colorado's MTSS model had six key components: (a) shared leadership; (b) data-based problem solving and decision making;

(c) layered continuum of supports; (d) evidence-based instruction, intervention, and assessment practices; (e) universal screening and progress monitoring; and (f) family, school, and community partnering. More recently, CDE (2016) leaders decided to blend the universal screening and progress monitoring component within the larger, systems-level framework. As a result, the current Colorado MTSS model is comprised of five components: (a) shared leadership; (b) data-based problem solving and decision making; (c) layered continuum of supports; (d) evidence-based instruction, intervention, and assessment practices; and (e) family, school, and community partnering (CDE, 2016).

Multi-tiered system of supports implementation perception survey variables. The MTSS-IPS (Pierce et al., n.d.) is a 50-item instrument and most of the items directly or indirectly relate to each MTSS component identified by CDE (2014). For example, a survey item that relates to both shared leadership and data-based problem solving and decision making reads, "The leaders at my school provide clear expectations for the use of problem-solving based on student data." A second example of an item that ties both evidence based instruction, intervention, and assessment practices to family, school, and community partnering reads, "The staff at my school works collaboratively to use data to assess and support their peers for continuous improvement of instructional practices." Most of the MTSS-IPS items were constructed to address the MTSS components identified by CDE. Participants were also asked to share basic demographic information, the professional role they held during the 2013-14 academic year, and their years of education-related experience. The MTSS-IPS used two different Likert scales, which allowed participants to measure the MTSS implementation efforts in their individual settings. One Likert scale was used for each item. The first was a numerical frequency scale ranging between 1 and 5 (1 = *Never*, 2 =*Rarely*, 3 = *Sometimes*, 4 = *Frequently*, 5 = *Almost Always*) with a sixth option being *I don't know.* The second scale used in the MTSS-IPS was used by survey responders to measure their MTSS implementation perceptions within their individual setting (1= *Not Evident*, 2= *Beginning to Be Established,* 3= *Partly Established*, 4= *Mostly Established,* and 5= *Fully Established*). The second scale also provided participants with a sixth option allowing them to indicate they had a lack of knowledge about the item which impeded them from providing useful information (6 = *I don't know*).

The number of survey items forced to load on each of the above latent factors varied by factor. For example, responses from 14 survey items were forced to load on the team-driven shared leadership (TDSL) latent factor, responses from five survey items were forced to load onto the EBP factor, and three survey items were forced to load onto the layered continuum of supports (LCS) factor. However, each latent factor, regardless of the number of survey items it contained, contributed equally to the overall model. A more detailed explanation of the individual survey items by individual component follows.

Multi-tiered system of supports implementation perception survey component I: Team-driven shared leadership. According to CDE (2016), TDSL is defined as the structures and expectations that spread responsibility and shared decision-making at the school, district, and community levels to shape synchronized systems of training, coaching, resources, implementation, and evaluation of adult activities. The MTSS-IPS survey items forced to load on the TDSL latent factor are presented in Table 1.

Table 1

Multi-Tiered System of Supports Implementation Perception Survey: Team-Driven Shared Leadership Items

Item Number	Item
1	The leaders and staff at my school collectively examine practices and processes of a multi-tiered system of supports frequently enough to ensure that they are improving outcomes for all students.
2	The leaders at my school provide clear expectations for the use of problem-solving based on student data.
3	The leaders at my school provide clear expectations that full implementation of a multi-tiered system of supports is necessary to improve the progress of all students.
4	The leaders at my school provide coaching and/or professional development opportunities to ensure that all staff members have the skills necessary to use data for problem solving.
5	The leaders at my school promote collaboration and trust among educators and families to meet the needs of students.
6	The leaders at my school request and welcome input from staff to revise school policies and procedures.
10	The climate at my school allows the staff and leaders to feel safe discussing school-related problems candidly.
11	My school has systematic processes that leaders utilize to ensure that staff has appropriate resources (e.g., personnel, time, materials) to implement a multi-tiered system of supports.
12	The leaders at my school work to ensure that the staff has a shared commitment for all students' learning and growth.
13	The staff at my school believes that full implementation of a multi-tiered system of supports is necessary to improve the progress of all students.
14	The leaders at my school model how to interpret and use student data for decision making.
15	The leaders at my school monitor the school's progress toward full implementation of a multi-tiered system of supports.
22	The leaders at my school actively participate in problem solving team meetings.
36	The staff and leaders at my school encourage a climate where families feel safe discussing their child's needs.

Multi-tiered system of supports implementation perception survey component II: Data-based problem solving and decision making. The CDE (2016) defined the data-based problem solving and decision making component as a consistent method and process used at numerous levels to examine and evaluate appropriate information to design and implement strategies that support increased student and systemic outcomes in a sustainable manner. The MTSS-IPS survey items forced to load on the DBPS latent factor are presented in Table 2.

Table 2

Multi-Tiered System of Supports Implementation Perception Survey: Data-Based Problem Solving and Decision Making Items

Item Number	Item
7	The staff at my school uses school-wide achievement trends to decide about interventions and/or instructional strategies for the following year.
8	The staff at my school analyzes the overall impact of student interventions at the targeted and intensive level at least annually to ensure that the interventions are effective.
9	My school follows a decision-making process that increases the frequency of progress monitoring as the intensity of instruction and intervention increases.
16	The staff engaged in problem solving processes at my school works to address the instructional needs of all children in the school, regardless of their academic level.
17	The staff engaged in problem solving at my school are collectively able to identify appropriate research-based interventions and instructional strategies for students at all academic levels.
18	The problem solving process at my school allows the staff to adjust instructional supports based on student data/results.
19	The staff engaged in problem solving at my school uses data to identify individual student need for targeted and intensive intervention.
20	The staff engaged in problem solving at my school uses data sources in addition to summative data from the state to analyze achievement trends collectively for all students.
21	The staff at my school use data to evaluate the effectiveness of our math curriculum.
23	The staff at my school use data to evaluate and improve their own instructional practices.
24	The staff at my school works collaboratively to use data to assess and support their peers for continuous improvement of instructional practices.
25	The staff at my school collects and analyzes information to determine whether differentiation of instruction occurs based on student need.
29	Defined decision-making processes at my school enable the staff to efficiently select interventions or instruction based on the level of student need.
32	Members of my school’s problem solving team have clear roles and responsibilities.
40	My school has a data management system for tracking academic progress of all students that is functional, useful, and accessible by all staff.
41	My school has a data management system to track school-wide behavior data (e.g., discipline referrals, truancy, attendance) that is functional, useful, and accessible by all staff.

Table 2 Continued

Item Number	Item
42	The staff at my school is proficient in accessing achievement data for our students.
43	The staff at my school knows how to interpret data to inform instructional practices.
44	The staff at my school uses standardized formative assessments (e.g., AIMSweb, Galileo, NWEA) to monitor student progress.
45	The staff at my school uses informal classroom formative assessments (e.g., observations, classroom quizzes, exit tickets, walk-arounds) to identify the immediate instructional needs of our students.
46	The staff at my school uses universal screening measures to identify any students needing additional supports to progress from their current academic level (e.g., accelerated, delayed, etc.).
47	My school administers universal screening and benchmarking assessments in math at regular intervals.
48	My school's assessment system provides guidelines on types of data needed to establish a body of evidence for eligibility for gifted services.
49	My school’s assessment system provides guidelines on types of data needed to establish a body of evidence for eligibility for all categories of special education.
50	All students at my school are involved in monitoring their own progress for the purpose of setting their own academic goals.

Multi-tiered system of supports implementation perception survey component III: Family, school, and community partnering. Family, school, and community partnering (FSCP) is defined by CDE (2016) as “the collaboration of families, schools, and communities as active partners in improving learner, classroom, school, district, and state outcomes” (p. 1). The MTSS-IPS items that aligned with FSCP are provided in Table 3.

Table 3

Multi-Tiered System of Supports Implementation Perception Survey: Family, School, and Community Partnering Items

Item Number	Item
37	The staff at my school increases interactions with parents as a student's needs increase.
38	The staff at my school engages families in conversations about student performance data, at least during each parent-teacher conference.
39	My school helps families understand student performance data for meaningful conversations about student progress.

Multi-tiered system of supports implementation perception survey component IV: Layered continuum of supports. Within the MTSS framework, a layered continuum of supports ensures all students are provided with academic and behavioral supports that are culturally responsive, individualized, and developmentally appropriate in increasingly intensive levels of support that move from the universal tier (every student) to the targeted tier (some students) to the most intensive (few students; CDE, 2016). The MTSS-IPS survey items that primarily addressed the concepts associated with the layered continuum of supports are provided in Table 4.

Table 4

Multi-Tiered System of Supports Implementation Perception Survey: Layered Continuum of Supports Items

Item Number	Item
27	The staff at my school regularly meets to determine instructional grouping of students.
28	The curriculum at my school is flexible enough for staff to differentiate instruction based on the individual needs of students.
31	The staff at my school uses a continuum of increasingly intensive instruction based on student needs and performance levels: all students (universal), some students (targeted), and a few students (intensive).

Multi-tiered system of supports implementation perception survey component V: Evidence-based practices. Finally, Colorado (2016) defines evidence-based practices (EBP) as "approaches to instruction, intervention, and assessment that have been proven effective through research indicating improved outcomes for students" (p. 1). The MTSS-IPS survey items that addressed concepts associated with evidence-based practices are listed in Table 5.

Table 5

Multi-Tiered System of Supports Implementation Perception Survey: Evidence-Based Practices Items

Item Number	Item
26	My school makes a range of opportunities for coaching and professional development that are aligned to each teacher's specific needs readily available throughout the year.
30	The staff at my school has enough research-based instructional options available to meet the needs of all students.
33	The staff at my school explicitly teaches appropriate behaviors expected of students.
34	When students fail to show appropriate behavior, staff respond by reinforcing the behavioral expectations as they were taught.
35	The staff at my school engages in classroom management techniques which creates a positive learning environment for all students.

Content and construct validity. After the CDE (2016) item review, national RTI experts were contacted by the MTSS-IPS survey development team to review both the construct and content validity of the instrument. Four national experts provided input, three of whom were affiliated with the National Center on Response to Intervention, which receives federal funding from the Federal Office of Special Education Programs. Following the recommendations made by Hinkin (1998), these experts were asked to rate both the importance and value of each survey item and how well it could be used to measure MTSS implementation in educational settings. The feedback these experts provided was used to revise survey items. Finally, the survey development team also facilitated guided focus group discussions with a small sample (N = 28) of educators. Feedback and suggestions provided by the focus groups primarily focused on item quantity, clarity and interpretation. As in previous validation efforts, the survey development team incorporated focus group suggestions in a final round of item revisions. These revisions resulted in a 50-item survey that captured the most essential elements of the MTSS.

Population and Sample

The MTSS-IPS was originally distributed to a sample of schools in both Colorado and northwestern Nebraska. The survey development team employed a well-thought out methodology to distribute the MTSS-IPS survey instrument. According to CDE (2016), during the 2014-15 academic year, over 60,000 teachers and school- and district-level administrators either directly or indirectly served students in K-12 settings across the state. While some of these individuals either led MTSS implementation efforts at the school-level and/or worked directly with students in classrooms to provide tiered and differentiated instructional and behavioral supports offered within the MTSS framework, some individuals might not have been intimately familiar with the MTSS framework due to the novelty of the initiative in the state. Therefore, the survey team began by identifying individual Colorado school districts and individual schools within each of those districts. For small districts, defined as those with 10 schools or less, the survey was sent to every potential participant randomly identified by role using Qualtrics survey distribution software. For larger school districts, defined as those with more than 10 elementary schools, the MTSS-IPS was randomly distributed to a group of schools and to target individuals randomly identified within those schools. An additional criterion for these larger districts meant the MTSS-IPS survey was distributed to 10 schools or no less than 30% of the total schools within the district, whichever was more. Of the 178 school districts in Colorado, 21 were large enough to require random sampling. In all other cases, every elementary school was included.

Using the above methodology, the sample included approximately 1,500 individuals employed in approximately 500 public elementary schools in Colorado. Target individuals included those with the following roles: third grade classroom teachers, school principals, or special education resource teachers who were randomly identified within the same school by role. After two weeks had elapsed, if a target individual had not started the survey, an invitation was sent to a randomly identified proxy individual who had an identical or closely-related professional role within the same school. For example, when a target third grade teacher did not complete the survey, an alternative and randomly identified third grade teacher from the same school was invited to participate. In very small schools where a second individual with an identical role was not available, randomly identified proxy individuals who served in a similar role (e.g., second, fourth, first, or fifth grade classroom teachers) received an invitation to complete the survey.

Of the 518 individuals who completed the MTSS-IPS and worked in Colorado, 28.6% (n = 148) were proxies. Individuals with proxy roles were provided with the opportunity to share their professional role at the end of the survey. Broadly speaking, the professional roles of this small subgroup of proxy individuals fell into one of three categories. Approximately 28% (n = 11) were school psychologists or counselors. Individuals with a reading-related role (e.g., reading specialists, Title I teachers, or reading interventionists) made up 26% (n = 10) of the subgroup. Finally, approximately 44% (n = 17) of the subgroup had an administrator-type role within their school (e.g., assistant principal, dean of students, or district superintendent). By the end of the survey distribution window, 518 individuals (target and proxy) had completed at least one question. Therefore, a conservative total response-rate estimate for the initial MTSS-IPS distribution was 25.06%.

Instrument Reliability Estimates

Table 6 provides the overall mean, minimum, maximum, and variance of the 50-item MTSS-IPS after any row with missing data was deleted.

Table 6

Colorado Summary Item Statistics

Mean	Minimum	Maximum	Variance
3.869	3.215	4.451	.087

Reliability measures the extent to which an item, scale, or instrument yields the same score when administered in different times, locations, or with different individuals (Groves et al., 2009). More precisely, initial reliability estimates provide reasonable proof of technical adequacy of the instrument. Cronbach's alpha (α; Cronbach, 1951) is currently the most common form of reliability coefficient. Scores for α range between zero and one (Remler & Van Ryzin, 2015). The model used to calculate α is represented by the following formula:

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

The reliability of the statistic α depends on the number of items, represented by k, and their average intercorrelation, represented by $\bar{r}$. A high value of Cronbach's alpha implies either high reliability or low response variance. Unfortunately, it can also indicate the answers to one item affected the responses of other item(s), which in turn induced high positive correlations. Conversely, a low α-value could either indicate low reliability or demonstrate the items did not really measure the same construct (Groves et al., 2009). Common rules of thumb of α-levels for internal consistency suggested by Groves et al. (2009) are displayed in Table 7.

Table 7

Internal Consistency Statistics

Cronbach's alpha	Internal consistency
$\alpha \geq .9$	Excellent
$.9 > \alpha \geq .8$	Good
$.8 > \alpha \geq .7$	Acceptable
$.7 > \alpha \geq .6$	Questionable
$.6 > \alpha \geq .5$	Poor
$\alpha > .5$	Unacceptable

Typically, α-levels should be .70 or higher to keep an item in a scale. Overall Cronbach's α for the MTSS-IPS was .97. Using the information provided in Table 7, one would expect the instrument provides exceptionally consistent data. However, it should be noted that very high reliabilities (.95 or higher) are not necessarily desirable as this indicates some of the items on the measure might be unneeded and redundant (Streiner, 2003). Initial α levels for the MTSS-IPS (.98) indicated the measure might benefit from item reduction efforts. Conversely, high α values might simply be due to the large number of survey items included in the MTSS-IPS survey. Therefore, reliability estimates were also calculated by MTSS components using SPSS software (International Business Machines Corporation [IBM], 2016). The reader might wish to refer to Tables 1-5 for item/component alignment. Component-specific reliability estimates for the MTSS-IPS are provided in Table 8.

Table 8

Multi-Tiered System of Supports Implementation Perception Survey: Cronbach's Alpha Reliability Estimates by Component

Component	α	Valid Cases
TDSL (14 items)	.96	275
DBPS (25 items)	.96	233
FSCP (3 items)	.81	290
LCS (3 items)	.74	293
EBP (5 items)	.84	292

Note. α = Cronbach's Alpha; Valid Cases = Valid After Listwise Deletion; TDSL =Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; FSCP = Family, School, and Community Partnerships; LCS= Layered Continuum of Supports; EBP = Evidence-Based Practices

Student Reading Achievement

When the NCLB Act of 2001 was passed, states were required to report the educational skills and progress of all children enrolled in a public school (Hudson, 2013). In Colorado, student achievement has been measured by schools and districts using an assortment of tools that have varied over time and by purpose. For example, Measures of

Academic Progress (MAP), a tool created by the Northwest Evaluation Association (NWEA; 2017), is primarily used to measure student growth in reading, language, science, and math over time. More specifically, the MAP assessment is a grade-independent, computer-adaptive, K-12 assessment that automatically adjusts to each student and his/her unique instructional level. To explain, when a student taking a MAP test answers a question correctly, the next question is more demanding. However, if the student answers a question incorrectly, the next question is less demanding (NWEA, 2017). By adapting the level of difficulty on a question-by-question basis, MAP accurately measures every student's individual level of achievement and growth over time, which is then used to meet students' needs at school- and classroom-levels (NWEA, 2017).

While the reliable and valid data the MAP assessment generates would provide an ideal source of information of elementary students' reading achievement, MAP data are not publicly available. Instead, MAP data are provided to individual schools within local districts. To incorporate this measure into the design of the study would have required approval at individual district- and school-levels, which would have significantly impeded the research process.

Fortunately, data obtained using an alternative measure was publicly available that assessed student achievement against a set of predetermined Colorado-specific standards called the Transitional Colorado Assessment Program (TCAP). According to CDE (2014), the TCAP is a standards-based assessment used to measure student performance across a variety of content areas including reading, writing, math, and science of Colorado students in grades 3-12 from 2012-2014. The primary purpose of the

TCAP measure was to generate data to determine the level at which Colorado students were able to meet the newly revised Colorado Academic Standards (CAS)--10 content areas standards aligned with the Common Core State Standards for reading, mathematics, writing, communicating that emphasize the skills and knowledge students need to be postsecondary and workforce ready. Additionally, the CAS incorporates a variety of 21st century skills in literacy, collaboration, critical thinking, self-direction, and invention. Further, the CAS were written with skill mastery as the expectation, which is defined as a student's ability to fluently apply and transfer knowledge and skills from across content areas and settings. Literacy skills were a key component and focus during the TCAP development process, as the CDE believed literacy forms the foundation of academic success for all students.

While in use, the TCAP (CDE, 2014) was intended to provide a variety of educational stakeholders (e.g., parents and school-, district-, and state-level administrators and policy makers) with school-, district- and state-level achievement results; it was collaboratively developed by CDE, members of the Colorado teaching community, and employees of CTB/McGraw-Hill (CDE, 2014). According to state law, every student enrolled in a public school was required to take the TCAP (or an alternative) for the appropriate grades and across the content areas. The TCAP was designed to certify that all districts serving students in Colorado were held to the same standards irrespective of whether students lived in urban, suburban, or rural areas. Student scores for each of the content area assessments were placed along a categorical continuum ranging from "Unsatisfactory," "Partially-Proficient," "Proficient," and "Advanced." The TCAP was used to generate data that were analyzed to evaluate students' mastery and growth of the CAS and to evaluate the performance of districts and individual schools. While the TCAP measure was used to assess students' literacy-related skills and competencies between 2012 and 2014, school-level reading data were publicly available and provided a solid alternative measure to quantify student reading achievement; data for small schools and/or districts with low student counts (*n* < 16) were not publicly reported in order to maintain student confidentiality. Because the MTSS-IPS asked participants to specifically answer survey items focusing on the 2013-2014 academic year, third grade 2014 TCAP data were used as the independent variable in the model. Third grade student achievement was specifically isolated because previous research findings indicated that after third grade, students transition from learning how to read and begin reading to learn (e.g., Al Otaiba et al., 2014; Wanzek & Vaughn, 2007). State-level third grade participation rates, proficiency percentages, and the percentage of students who did not receive a score because they opted out of taking the test, were absent, or because their test was not scored due to a misadministration error are included in Table 9.

Table 9

Third-Grade Transitional Colorado Assessment Program Results for 2014

Total	U		PP		P		A		NS	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
63,665	6447	10.13	11326	17.79	41376	64.99	4196	6.59	320	.5

Note. U= Unsatisfactory; PP= Partially Proficient; P= Proficient;
A= Advanced; NS= No Score; *n* = number; % = Percentage of third grade students in the category

Structural Equation Modeling

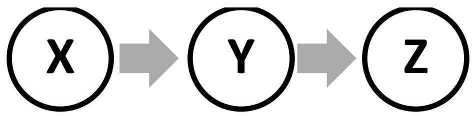
Having established that the MTSS-IPS was a valid instrument capable of providing reliable data, a description data analysis process followed. Structural equation modeling (SEM) was used as the main analytic approach for this study. To provide the reader with a basic understanding of the concepts, advantages, and methods of SEM, a brief overview of the method is provided.

According to Kline (2016), SEM is not a single statistical method or technique. Rather, SEM is a family of procedures rooted in the foundation of regression analysis of observed variables and in factor analyses of latent variables that researchers can employ as they attempt to investigate and identify causal inferences. The basic features of SEM incorporate a priori structural model creation and analyze covariance matrices of that model to assess how well data fit the hypothesized model (Kline, 2016). Like regression analysis techniques, SEM is significantly influenced by the variables included and measured in the model as well as those omitted (Kline, 2016). In SEM, omitted predictors that tend to covary with included predictor variables might induce errors in analysis. Therefore, researchers using SEM should strive to identify and choose predictor variables only after completing a thorough review of related research literature.

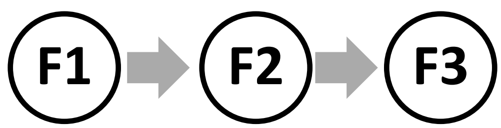
According to Kaplan (2000), SEM blends confirmatory factor analysis (CFA) with path analysis (PA) and provides a statistical model that researchers can use to examine complex interactions between both measured and unobservable or latent variables. Lei and Wu (2007) noted one of the principal advantages of SEM when compared to other models based on general linear models (e.g., t-tests, F-tests, and assorted analysis of variances and covariance) was the SEM could be used to study relationships among latent constructs that have been measured by multiple measures and could be used to analyze experimental, quasi-experimental, non-experimental, cross-sectional, and longitudinal data. Similarly, Ullman (2006) shared that SEM methods permit researchers to identify the relationships between one or more independent (predictor) variables with one or more dependent (criterion) variables.

Like many statistical terms, SEM has a variety of labels: covariance structural modeling, latent variable modeling, and causal modeling (Ullman, 2006). As one might expect, the label “causal modeling” provokes researchers to criticize SEM for violating the hallowed statement made by statisticians the world round: “Correlation does not imply causation” (Pearl, 2009, p. 99). However, SEM uses the term “causal” to describe the relationship between variables identified by the researcher before any data have been gathered. Only after researchers have reviewed the body of literature and combined the knowledge gained as a result of that review with systematic deliberation are they able to make deductive inferences about the direction of causality between the variables. As a result, simple correlations between a set of variables in a SEM should not be used to proclaim one variable causes another. Instead, if the variables are causally related, the basis for that relationship lies in the direction of the relationship instead of the opposite (Keith, 1999). According to Byrne (1998), researchers using SEM impose a structure on the data (e.g., force the data to fit the theorized model) and then decide how well the observed data fit that model in an iterative manner. How well the data fit the hypothesized model is the residual term in SEM. The ultimate aim of SEM is to increase the strength of research findings by increasing the accuracy of the relationships that exist between the theoretical constructs in a hypothetical model (Kline, 2016).

In many models of statistical analysis, independent variables (IV) and dependent variables (DV) can be continuous, dichotomous, or ordinal. In SEM, both IVs and DVs can be either measured or unmeasured (Ullman, 2006). As previously noted, SEM can be viewed as a combination of path analysis and CFA. The focus of path analysis is to order measured variables a priori (Kline, 2016). As an example, the reader might want to consider the following example of a path analysis model:



In the above model, X, Y, and Z are measured variables and the arrows are used to represent hypothesized causal effects of the model. In other words, X leads to Y, which in turn leads to Z. A full structural equation model is similar to a path model but the focus shifts from finding the appropriate order of variables to ordering the latent (unobserved) factors a priori. In a SEM, factors (represented in the diagram below by F1, F2, and F3) represent the latent (unmeasured) variables.



The latent factors of a structural model are used to identify the direction of causal relationships (Klem, 2000). In the above example, the first factor (F1) causes the second factor (F2), which in turn causes the third factor (F3). The relationship between the three factors is the SEM.

In SEM, the term "model" describes the (a) measurement, (b) structural, or (c) full model. Structural equation modeling begins by testing the fit of the measurement model. Unlike traditional path analysis, which both assumes perfect measurement of and between the variables and does not include any measurement error, SEM includes estimations of the model's measurement error or residuals. To evaluate the measurement model, CFA techniques are used a priori to test the hypotheses between the latent (unmeasured) factors and the observed indicators using covariance and error terms. Using CFA techniques provides the researcher with the opportunity to identify specific factor loadings and correlate both residuals and factors.

The second component of SEM is the structural model, which examines the displays the internal paths between the proposed model and the theoretical relationships between the model's latent factors and their observed indicators. According to Schreiber, Nora, Stage, Barlow, and King (2006), when examining the structural model, path coefficients are first estimated and subsequently used to analyze the direct, indirect, and total effects of the model. One key difference between path modeling and SEM is in path analysis it is assumed a single indicator (IV) perfectly measures the latent factor; while in SEM, multiple indicators are used to estimate each latent variable, assuming a single indicator is usually not capable of perfectly measuring and estimating a latent construct. In SEM, if the a priori model is correctly specified, it is "identified." However, if the data do not fit the model, the researcher moves on to test a different, re-specified model.

Model Evaluation and the Role of Fit Indices

In SEM, global fit statistics are used by the researcher to decide whether an a priori model adequately fits the data. If data do provide adequate fit to the hypothesized model, then the model is retained and identified. Because so much hinges on the fit of the data in SEM, it should not be surprising there is little agreement in the statistical community about the principles and rules researchers should adopt when interpreting global fit indices. While a variety of guidelines can be used while interpreting fit statistics, Kline (2016) cautioned these statistical rules of thumb should not be applied in a universal way and should never take the place of sound judgement rooted in solid theory on the part of the researcher.

Literally dozens of fit indices have been used to evaluate model fit and more are being developed all the time. Whether novel or traditional, all global fit statistics have limitations (Kline, 2016). Researchers using SEM methods should always remember that global fit statistics only specify average model fit because they condense many discrepancies into a single number (Steiger, 2007). However, researchers should also remember when using SEM, "there is no such thing as a magical, single number summary that says everything that is worth knowing about the model fit" (Kline, 2016, p. 264). Additionally, Millsap (2007) shared global fit statistical values do not provide any information that can be meaningfully used to assess models with poorly fitting data. In other words, global fit index statistics do not provide any useful information on the direction of relationships or whether the researcher has specified the correct number of factors in the CFA. Further, global fit statistic values that indicate adequate fit should not be used in the same way as the R^2 effect value. Kline (2016) cautioned the R^2 value and overall model fit values are independent of each other and shared that fit statistics are not capable of assigning meaning to a poorly specified model that lacks a strong theoretical foundation. Finally, because global fit statistics only assess average model fit, they should not be used to examine person-level fit, which is the extent the model produces accurate estimates for individual cases (Kline, 2016). Having developed a basic understanding of what global fit indices are capable of, a discussion of some of the potential fit indices that might be used to evaluate the fit of the proposed model follows.

As noted, in SEM, a researcher's primary focus is to determine if a set of data provides adequate levels of fit to the hypothesized model(s). Kline (2016) recommends including the following fit statistics when reporting SEM results: (a) the maximum likelihood chi-square statistic (χ^2_{ml}), (b) the root mean square error of approximation (RMSEA; Steiger, 1990), (c) the Bentler comparative fit index (CFI; Bentler, 1990), and (d) Tucker Lewis index (TLI; Tucker & Lewis, 1973). An overview of each statistic follows.

A common statistic used to evaluate the overall global fit of the data to the model is the maximum likelihood ratio chi-square test (χ^2_{ml}), which is used to evaluate whether a difference exists between the model's sample population covariance predictions and those of the larger population (Hooper, Coughlan, & Mullen, 2008). If $\chi^2_{\mathrm{ml}} = 0$, the model perfectly fits the data, meaning the observed sample covariance is statistically similar to its proposed population covariance equivalent. When a model is incorrectly specified, the χ^2_{ml} value increases. Therefore, a small χ^2_{ml} is preferred. While commonly reported in SEM research, the χ^2_{ml} tends to be affected by multivariate non-normal data, significant correlations among the observed variables, and large sample sizes (Kline, 2016). Additionally, larger sample sizes produce larger values for the χ^2_{ml} statistic and increase the probability of a Type I error (Schreiber et al., 2006). Given sample size should never be a factor that affects the fit of the model to the data, researchers are cautioned to interpret the χ^2_{ml} carefully and neither respecify or accept the model without examining additional fit indices.

Developed by Steiger and Lind (1980), the RMSEA index is based on the idea that because no model will ever perfectly match that of the population, the best result a researcher can ever hope for is one that provides a close approximation of that ideal. More specifically, the RMSEA is an absolute fit index with a scale ranging between zero and one. A value close to zero indicates the best outcome and result. The RMSEA is sensitive to the number of parameters in the model but unlike the χ^2_{ml}, it is not sensitive to sample size. For RMSEA, four levels of model fit are suggested: (a) values less than .05 indicate good model fit, (b) values ranging between .05 and less than .08 imply reasonable model fit, (c) values ranging between .08 and .10 indicate mediocre fit, and (d) values greater than .10 indicate poor fit (Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996).

The two final indices recommended by Kline (2016) are incremental fit indices that assess chi-square value to the baseline model. Originally proposed by Bentler (1990), the CFI is one of the most reported fit indices in SEM literature (Schreiber et al., 2006). Similar to the RMSEA fit index, values of CFI goodness-of-fit index range between zero and one. Comparative fit index values close to one indicate a better fit of the data to the model. Values for the TLI also range between zero and one; larger values also indicate a better fit of the data to the theorized model (Schreiber et al., 2006).

These fit indices were generated using the weighted least square means and variance (WLSMV) estimator (Beauducel & Herzberg, 2006). Briefly, the WLSMV estimator does not require large sample sizes, handles the intricacies of categorical data (e.g., Likert-scale items), and works with measures with large variables. For these reasons, the WLSMV estimator was the ideal estimator and was employed during the data analysis phase of this study.

Steps to Structural Model Identification

A researcher uses a series of basic steps in most SEM analyses (Kline, 2016). Figure 4 provides a visual representation of those steps in order for the SEM novice to understand the model. As a reminder, the SEM models that guided this research were included in Figures 2 and 3.

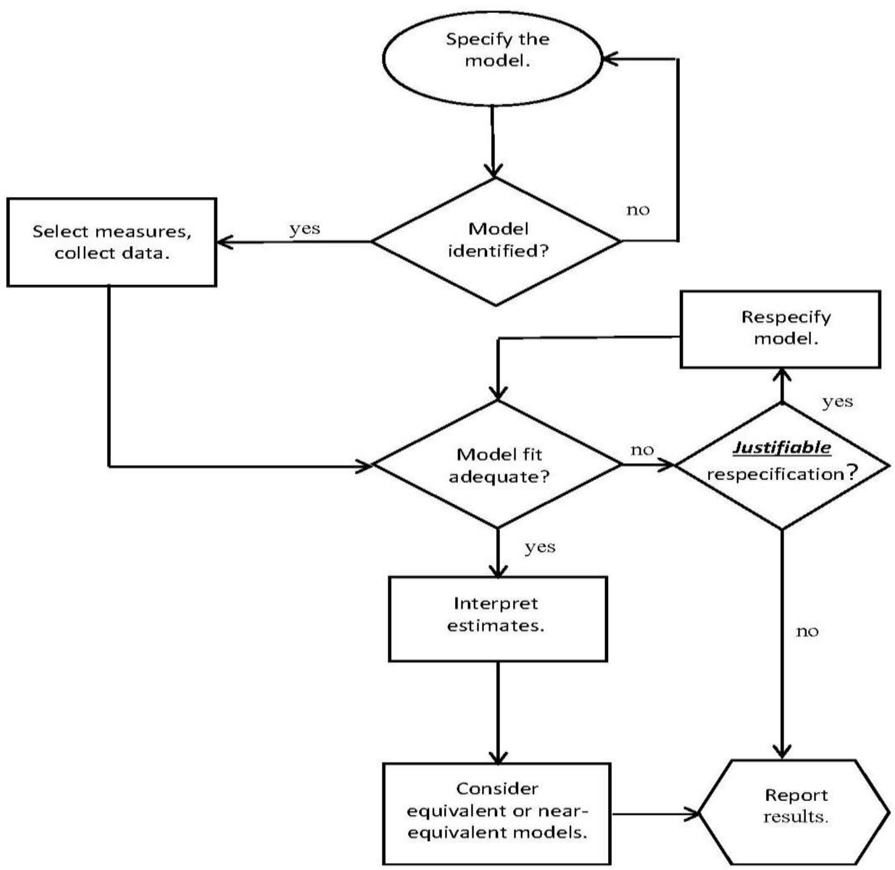


Figure 4. Basic steps of structural equation modeling. A defensible model respecification has a basis in theory and/or previous research findings. Respecification is only warranted when the respecified model is identified a priori. Adapted from *Principles and Practice of Structural Equation Modeling* (p. 118) by Kline (2016).

Methods for Handling
Missing Data

In most methods of statistical analysis, analyzing complete sets of data without any missing data is the ideal. However, in the real world, missing observations appear in many, if not most, data sets (Kline, 2016). While a small percentage of missingness (e.g., < 5%) might not cause a researcher to lose any sleep, larger levels of missing observations are more challenging, especially if the data are not missing simply by chance. Some of the traditional methods historically used to account for missing observations in data sets include both listwise and pairwise deletion as well as a variety of imputation methods (e.g., mean imputation, group-mean substitution, or regression substitution). Unfortunately, these more traditional methods tend to yield biased results (Peters & Enders, 2002). While these more traditional methods are common, more modern and contemporary methods improve the level of bias in data. Briefly, model-based methods like full information maximum likelihood (FIML) start with the model, divide the raw data cases into subsets with similar patterns of missingness, calculate means and variances from each subset, and thus preserve all the cases for inclusion in the analysis (Allison, 2003). These estimates are included in the model's parameters once all information is combined across all subsets of cases. Many SEM software programs (e.g., M PLUS, LISREL/PRELIS) have included FIML as an available option for handling missing data (Kline, 2016). After receiving Institutional Review Board approval (see Appendix C), FIML methods were employed to examine how perceptions of MTSS implementation correlated with 2014 third grade TCAP reading scores.

CHAPTER IV

RESULTS

The purpose of this study was to investigate the relationship between third grade student reading achievement and educators' perceptions of MTSS implementation in Colorado. The MTSS framework is a new educational initiative that combines student-level academic supports of the RTI framework with the behavioral supports of the PBIS system into one systemic model focused on providing teachers with assistance and support as they strive to meet the needs of the students in their classroom. The previous chapters of this dissertation (a) established the MTSS framework is one that is defined in a variety of ways across the nation and (b) reviewed the research literature that examined individual elements of the MTSS framework. When possible, those reviews included studies that investigated MTSS-related elements and how they impacted student reading outcomes. Finally, a description of measures used to gather data utilized in this study was provided.

This study used existing survey data from the MTSS-IPS (Pierce et al., n.d.)--a survey instrument with 50 Likert-scaled, positively worded items that directly or indirectly related to each of the MTSS components identified by CDE (2016). The survey used two different 5-point Likert scales depending on the type of item presented. The first was a numerical frequency scale with response options ranging between 1 and 5 (1 = *Never*, 2 = *Rarely*, 3 = *Sometimes*, 4 = *Frequently*, 5 = *Almost Always*). The second scale provided participants with the opportunity to evaluate the level of MTSS implementation within their individual schools (1= *Not Evident*, 2= *Beginning to Be Established*, 3= *Partly Established*, 4= *Mostly Established*, and 5= *Fully Established*). Both scales included a sixth option responders used when they did not have enough information to provide meaningful information (6 =*I don't know*), which were treated as missing data during the analysis process.

Data collected from the MTSS-IPS were analyzed using a variety of software packages. Both IBM SPSS Statistics 24.0 (2016) and Mplus Version 8 (Muthen & Muthen, 2015) software programs were used. This chapter begins by sharing results of the preliminary analyses. These results describe the demographic information (e.g., gender, years of experience and within school) of the Colorado participants and share item-level descriptive statistics (mean, standard deviation, skewness, and kurtosis) for each indicator variable of the MTSS-IPS. After these results are reported, the results from each of the research questions are presented.

Sample

The major purpose of this study was to investigate how perception of MTSS implementation during the 2013-2014 academic years in Colorado related to 2014 third grade reading TCAP results. The survey instrument (the MTSS-IPS; Pierce et al., n.d.) was distributed to a variety of individuals working in elementary schools both in western Nebraska and throughout the state of Colorado. Because this study specifically focused on the relationship between Colorado perceptions of MTSS implementation and 2014 third grade reading TCAP scores, only responses from the Colorado sample were used.

The MTSS-IPS was initially distributed by the survey development team in November of 2014 to individuals who worked in approximately 500 Colorado schools with three specific professional roles: third grade teachers, special education resource teachers, and school principals. These target individuals were randomly identified at the school-level. When the first person on each of the target-role lists did not complete the survey within 15 days, the survey was sent to a second individual with an identical or closely similar (i.e., proxy) role. A total of 518 individuals from the Colorado sample responded to at least one question. However, many of the 518 responders only answered the first question that asked which state they were employed in before dropping out of the study as a participant. From the schools included in the sample, survey responses that provided more than simple state of employment information were provided by 376 individuals who worked in one of 306 schools, which yielded a conservative response rate of 25.06%. Moreover, MTSS-IPS responses were provided by individuals with either target or proxy roles who worked in schools in 133 of Colorado's 183 school districts. In sum, MTSS-IPS data were received from individuals in the vast majority (72.68%) of Colorado school districts.

On average, of survey responders who completed the demographic questions at the end of the survey, 86.47% were women (*n* = 243) and 13.52% were men (*n* = 38). Approximately 62% of the responders reported they did not have a special education license or endorsement (*n* = 178). As of 2014, 91.90% of the responders had been a licensed educator for at least four years (*n* = 261) and 70.11% of that total reported they had been a licensed for more than 10 years (*n* =183). Of the total, 12.75% of the responders were principals (*n* = 37), 27.58% (*n* = 80) were special education resource teachers, 38.96% (*n* = 113) were third grade classroom teachers, and 20.68% (*n* = 60) had a proxy role (e.g., assistant principal, counselor, dean of students, interventionist, or school psychologist). Of the responders, 93.77% had served students in the same capacity during the 2013-2014 academic year (*n* = 271). When asked to share how many years they had been an educator at their current school, 68.66% of the responders (*n* = 195) had worked in the same setting more than three years. Table 10 provides a summary of the participants' demographic characteristics.

Table 10

Demographic Characteristics of Participants

Variables	*n*	%
Gender (*n* = 281)		
Male	38	13.52
Female	243	86.47
Years as Licensed Educator (*n* = 284)		
< 1 year	5	1.76
1-3 years	18	6.33
4-10 years	78	27.46
>10 years	183	64.43
Current Position (*n* = 290)		
Principal	37	12.75
Special Education Resource Teacher	80	27.58
Third Grade Classroom Teacher	113	38.96
Other (Proxy)	60	20.68
Current Position in 2013/2014 (*n* = 289)		
Yes	271	93.77
No	18	6.23
Years in Current Position (*n* = 266)		
< 1 year	1	0.04
1-3 years	79	29.70
4-10 years	97	36.47
>10 years	89	33.46
Years at School (*n* = 284)		
< 1 year	14	4.92
1-3 years	75	26.40
4-10 years	102	35.92
>10 years	93	32.75
Special Education License or Endorsement (*n* = 287)		
Yes	109	37.98
No	178	62.02

Note. Information above was provided by MTSS-IPS Colorado participants; *n* = Number of Responders; % = Percent of Responders

Preliminary Analyses

Before analyzing the data using SEM techniques, several preliminary analyses were performed to determine if any statistical assumptions associated with SEM had been violated. These analyses were important because previous research had demonstrated that SEM is sensitive to data that significantly depart from the normal distribution and are highly collinear (Grewal, Cote, & Baumgartner, 2004; Kline, 2016). Therefore, these preliminary analyses were conducted to ensure the MTSS-IPS data from the Colorado sample were normally distributed. These descriptive analyses generated item-level means, standard deviations, skewness, and kurtosis values. In the narrative that follows, those values are reported at both the item and indicator levels. The final portion of this segment includes reliability estimates on scores based on the MTSS-IPS measure as a whole as well as on individual latent variables included in the structural model.

Before using SEM to assess the fit of the data to the proposed models, several preliminary data analyses were conducted to determine if the data were normally distributed because SEM is sensitive to non-normally distributed data (Kaplan, 2000). Byrne (1998) suggested skewness values ranging between +/- 1 and kurtosis values ranging between +2 and -1 are normally distributed. However, data sets with skewness values greater than +3 are typically considered extremely skewed, while an absolute value of kurtosis that exceeds 10 indicates more serious normality problems (Byrne, 1989; Kaplan, 2000). Each of the MTSS-IPS items' means, standard deviations, skewness, and kurtosis values was examined. Of the 50 MTSS-IPS items, 90.0% fell well within the acceptable range for skew and kurtosis (n = 45), while 10.0% were only mildly or moderately skewed or kurtotic (n = 5). Of the five, none fell into an extreme range. Therefore, no revisions were made to the small sample of MTSS-IPS items that were just outside the normal range for skewness and kurtosis. The results for each survey item including item-level means, standard deviations, and levels of skew and kurtosis well as values for the dependent variables and covariates are reported in Appendix D.

To assess potential non-response bias, a χ^2 test of independence was conducted between the target and proxy groups related to their years of experience as educators, gender, and years of experience working within a specific school. A significant difference between target and proxy individuals might have supported the idea that there were systematic differences between the respondents and non-respondents, which would pose a threat to the generalizability of the results (Groves & Peytcheva, 2008). The assumption that grounded this analysis to deal with nonresponse bias was proxy individuals are similar to non-responders. If there was no difference between target and proxy individuals, then nonresponse bias would not pose a significant threat to the generalizability of the results because outside factors were likely responsible for the percentage of nonresponse results (Groves & Peytcheva, 2008). However, if a difference between target and proxy individuals existed, then the generalizability of the study would have been limited. A total of 283 usable respondents (target and proxy) were included in this portion of the data analysis; 73.85% (n = 209) were target individuals (third grade teachers, special education resource teachers, or school principals) while 26.14% (n = 74) individuals had proxy roles (e.g., second/fourth grade teachers, Title I teachers, or assistant principals).

Results demonstrated no differences between the groups in terms of years of experience, gender, and number of years at the same school. Therefore, it can be concluded the target and proxy individuals were from the same population, that this study's respondents were similar to Colorado educators on gender and years of experience, and nonresponse bias was not a significant concern in this study. Results of the chi-square test of independence are provided in Table 11.

Table 11

Chi-Square Test for Target and Proxy Individuals

Variable	χ^2	*df*	*p*-value
Gender	1.06	2	.59
Years in Education	3.34	6	.74
Years at School	.60	3	.90

Note. χ^2 = Chi-Square Test of Independence; *df* = Degrees of Freedom.

Evaluation Methods

Schreiber et al. (2006) shared when researchers are interested in studying unobservable, latent factors (e.g., MTSS), they use observable indicators to make statistical inferences about the correlation between the observable and latent variables being studied. Three statistical methods could be utilized to examine those relationships: exploratory and confirmatory factor analysis and structural equation modeling (SEM). These three statistical models reduce observable variables into a smaller set of latent variables using the covariation between the observable indicators (Schreiber et al., 2006). Confirmatory factor analysis (CFA) is a theory-driven technique that, as the name implies, confirms hypotheses about the theoretical relationships between the observed and unobserved variables of a hypothesized measurement model (Kline, 2016; Schreiber et al., 2006). When conducting a CFA, a hypothesized model is used to estimate a population covariance matrix that is then compared with an observed covariance matrix. Structural equation modeling is a statistical model that builds on the results supplied by CFA techniques and uses them to estimate the path coefficients by running a series of multiple regressions (Kline, 2016; Ullman, 2006). In other words, SEM uses the measurement component of CFA, creates path coefficients, and incorporates that information into a full structural model.

In SEM, variables can be observed or unobserved. Observed variables, or indicators, are graphically represented with a square or rectangle. Participants' responses to a 5-point Likert-scaled survey item are an example of an observed variable. Conversely, unobserved variables, or latent factors, are graphically represented with a circle (Schreiber et al., 2006). The components of MTSS (e.g., team-driven shared leadership; family, school, and community partnering) and a model's measurement errors are examples of unobserved variables. Arrows that point from a latent factor to an observed indicator are used to indicate a theorized causal effect of the latent factor on the observed indicator (Schreiber et al., 2006).

In SEM, unobservable and observable variables included in the model can be exogenous, endogenous, or both (Kline, 2016). Exogenous variables are roughly equivalent to independent variables, while endogenous variables are comparable to dependent variables. However, both endogenous and exogenous variables can be either observed or latent depending on the specific theoretical model being tested (Schreiber et al., 2006). Within an SEM, exogenous variables have an influence on other variables (Kline, 2016). Conversely, endogenous variables are influenced by other variables in the model (Schreiber et al., 2006). To summarize, SEM provides researchers with a statistical tool they can use to investigate theories about how latent constructs are theoretically linked and provides information that can identify the directionality of those relationships (Kline, 2016; Schreiber et al., 2006; Ullman, 2006).

Factor Analyses

Research question one asked whether the data provided an adequate fit to the proposed MTSS models hypothesizing relationships between implementation perceptions of MTSS in Colorado and 2014 third grade TCAP data. Structural equation modeling was used to answer this research question. The iterative process used when analyzing data with SEM and performing the series of statistical analyses provides researchers with the opportunity to identify areas that might impede the analysis when the full models are examined in the final stage. According to Ulmer (2004), "The full model is only as good as the individual components of the model (p. 96)." Consequentially, a series of confirmatory factor analyses (CFAs) was conducted using Mplus software (Muthen & Muthen, 2015) to study the patterns of the observed indicator variables within the latent factors included in the hypothesized model. The results of the initial CFAs were then used to assess the level of interrelationships of the individual latent variables that combined to create the MTSS framework. More specifically, the CFAs were conducted to investigate how well the inter-item covariance matrix fit the single factor models for each of the MTSS component scales.

In this study, a variety of CFA measurement models for each of the current and previous individual Colorado MTSS components (latent factors) were tested during this initial phase. To review, the current Colorado MTSS framework included the following five components: (a) team-driven shared leadership (TDSL), (b) data-based problem solving and decision making (DBPS), (c) evidence-based practices (EBP), (d) layered continuum of supports (LCS), and (e) family, community, and school partnerships (FSCP). This model was developed by current CDE (2016) leadership in an effort to create a more efficient MTSS framework. The previous MTSS framework included six latent factors: (a) team-driven shared leadership (TDSL); (b) universal screening and progress monitoring (USPM); (c) data-based problem solving and decision making (DBPS); (d) evidence-based practices (EBP); (e) layered continuum of supports (LCS); and (f) family, community, and school partnerships (FSCP).

The primary difference between the five-factor MTSS model and the previous six-factor model is the current DBPS factor includes a variety of data-related concepts that include data-gathering mechanism and processes and the ways those data are used by school-level teams. Specifically, the current expanded DBPS factor combines concepts associated with data gathering mechanisms formerly isolated in the USPM factor with the ways educators use data during instructional planning processes formerly encapsulated within the DBPS factor. A more detailed explanation of each of the MTSS latent factors was presented in Chapter III. After the fit of the MTSS-IPS data to the theorized models was evaluated, three additional CFAs were used to examine the overall fit of the MTSS-IPS data to a variety of theoretical MTSS models with a range of latent factors. The theoretical bases for the MTSS models are discussed in detail in later sections.

To evaluate the results of each of the CFAs, the following process was implemented. First, parameter specifications for the model being examined were inspected. The parameter specifications characterized the proposed relationships from the underlying latent construct (e.g., FSCP, LCS, and TDSL) to the observed MTSS-IPS survey-item indicators. The parameter estimates were inspected for size and directionality (i.e., positive or negative). Next, the values of the squared multiple correlations (SMC), a statistic that estimated how well an individual item measured an underlying latent construct with standardized values that range between zero and one, were inspected. Specifically, the SMCs were examined to establish and confirm the reliability of individual MTSS-IPS indicator variables. An SMC value less than .20 indicated the item might generate unreliable data, while values closer to one indicated higher reliability levels (Bollen, 1989; Ullman, 2006). Next, the completely standardized factor values, which estimated the correlation between the observed variables and the individual latent factor of interest, were inspected. These standardized values, henceforth represented by the symbol λ, were examined to ensure the value was statistically significant and fell into an appropriate range. Like the SMCs, factor loading values ranged between zero and one; higher values indicated stronger correlations. Comrey and Lee (1992) suggested using a range of λ values when assessing model fit. Loadings in excess of .32 (with 10% overlapping variance) should be considered poor. Loadings in excess of .45 (with a 20% overlapping variance) should be considered fair, while loadings of .55 (with a 30% overlapping variance) are typically considered good. Factor loadings of .63 (with a 40% overlapping variance) are very good, and loadings of .71with 50% overlapping variance are typically considered to be very good. Finally, chi-square (χ^2) values and a variety of fit indices were reviewed to determine how well the observed data fit the proposed model.

According to Hooper et al. (2008), when assessing the initial fit of models, absolute fit indices could be used to help evaluate how well the observed data fit the hypothesized model. One commonly reported absolute fit statistic is the chi-square test of model fit (χ^2; Hu & Bentler, 1999). However, while commonly reported, it should be noted that values of the χ^2 test statistic increase and become statistically significant as a sample size increases (Bentler & Bonnett, 1980; Joreskog & Sorbom, 1993). For this reason, in SEM, the χ^2 test statistic is often called the "badness-of-fit" statistic (Hu & Bentler, 1999). Commonly speaking, the χ^2 value "travels" with sample size; as sample size increases, so does the value for the χ^2 test statistic. Because SEM requires large sample sizes to produce reliable results, this global fit index is typically both large and statistically significant, which typically inaccurately indicates a poor fit of the data to the hypothesized model. Fortunately, other incremental fit indexes do not have the same propensities. For example, the root mean square error of approximation (RMSEA; Hu & Bentler, 1999) statistic has a range of zero to one and indicates a good fit of the proposed model to observed categorical data when its value is $< .06$ (Schreiber et al., 2006). Conversely, values for the CFI (Hu & Bentler, 1999) and the normed TLI (Hu & Bentler, 1999) also typically range between zero and one but larger values of these two statistics indicate a better fit of the observed data to the hypothesized model (Schreiber et al., 2006). The suggested cutoff level for an appropriate fitting model with categorical data is to have a CFI $> .95$ and a TLI $>.96$ (Schreiber et al., 2006). In general, if the majority of indices indicate a good fit, then there probably is a good fit of the data to the hypothesized model(s; Schreiber et al., 2006).

The survey responses provided by the Colorado sample were used to create a covariance matrix based on both polychoric correlations and the item-level standard deviations for all items of interest on the MTSS-IPS. A variety of models with a range of latent variables and a variety of exogenous variables were proposed and tested during this study. These models moved progressively from simple to complex. The final set of analyses examined a series of higher-order structural models that theorized a relationship between third grade TCAP reading scores, participants' perceptions of MTSS, district size, and/or students' socio-economic status.

Factor Analysis for Team-Driven Shared Leadership

The measurement scale of TDSL consisted of 14 items. The participants were asked to indicate their perceptions of shared leadership within their individual schools using a 5-point Likert scale. A CFA was conducted on the TDSL factor with 14 indicator variables. The results of the CFA for this measurement model were χ^2 (77, N = 382) = 434.93, p < .01; RMSEA = .11; CFI = .98; TLI = .98. The statistically significant value of the χ^2 statistic indicated the proposed model provided a poor fit for the observed data. However, as noted previously, the χ^2 statistic is sensitive to sample size and typically statistically significant in SEM. Therefore, other fit indices typically provide a more reasonable approximation of how well the observed data fit the proposed model. In this CFA, while the RMSEA value of .11 indicated the model provided a poor fit, the CFI value of .98 and the TLI value of .97 both indicated the model provided a good fit for the data. Similarly, the λ values ranged between .73 and .91, indicating all were statistically significant at the .01 level. The SMC values ranged between .53 and .82, which were well above the suggested minimum .20 value (Tabachnick & Fidell, 2007). These results, combined with the CFI and TLI fit indices, indicated a good fit of the MTSS-IPS data to the TDSL latent factor. Results of the CFA for the TDSL subscale are displayed in Table 12 (see Appendix E for the graphic representation).

Table 12

Confirmatory Factor Analysis for Team-Driven Shared Leadership

Indicator	λ	SMC
IPS1	.77*	.59*
IPS2	.84*	.70*
IPS3	.88*	.77*
IPS4	.75*	.56*
IPS5	.80*	.64*
IPS6	.77*	.60*
IPS10	.86*	.74*
IPS11	.83*	.69*
IPS12	.87*	.75*
ISP13	.83*	.69*
IPS14	.88*	.78*
IPS15	.91*	.82*
IPS22	.79*	.62*
IPS36	.73*	.53*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation
* $p < .01$.

Factor Analysis for Data-Based Problem Solving

As with the TDSL factor, the measurement scale of DBPS consisted of 14 individual items. The participants were asked to indicate their perceptions of the data-based problem-solving process within their schools using a 5-point Likert scale. Examples of the DBPS-related items included (a) the staff at my school uses school-wide achievement trends to decide about interventions and/or instructional strategies for the following year, (b) the staff at my school use data to evaluate and improve their own instructional practices, and (c) my school follows a decision-making process that increases the frequency of progress monitoring as the intensity of instruction and intervention increases (Pierce et al., n.d.). The results of the CFA were χ^2 (77, N = 377) = 548.98, $p < .01$; RMSEA = .13; CFI = .97; TLI = .96. As with the previous CFA, the χ^2 value was unsurprisingly statistically significant so other fit indices were evaluated. While the RMSEA value exceeded ideal values, both the CFI and TLI indicated satisfactory fit of the data to the estimated model. The λ values ranged between .70 and .91 and all were statistically significant at the .01 level. The SMC values for the 14 indicators ranged from .59 to .83. Overall, the CFA results indicated an adequate measurement model for the DBPS latent factor. Results of the CFA for the DBPS subscale are provided in Table 13 (see Appendix F for the graphic representation).

Table 13

Confirmatory Factor Analysis for Data-Based Problem Solving

Indicator	λ	SMC
IPS7	.79*	.63*
IPS8	.86*	.74*
IPS9	.85*	.73*
IPS16	.87*	.76*
IPS17	.89*	.79*
IPS18	.90*	.80*
IPS19	.91*	.83*
IPS20	.80*	.63*
IPS21	.73*	.53*
ISP23	.84*	.70*
IPS24	.85*	.72*
IPS25	.83*	.69*
IPS29	.85*	.72*
IPS32	.70*	.59*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation
* $p < .01$.

Factor Analysis for Universal Screening and Progress Monitoring

The USPM scale consisted of 11 individual items. Participants were asked to rate their school's perceptions of the universal screening and progress monitoring component of the MTSS framework using a 5-point Likert scale. Examples of USPM-related items included (a) My school has a data management system for tracking academic progress of all students that is functional, useful, and accessible by all staff; (b) My school administers universal screening and benchmarking assessments in math at regular intervals; and (c) The staff at my school uses standardized formative assessments (e.g., AIMSweb, Galileo, NWEA) to monitor student progress. A CFA was performed on this measurement model with 10 indicators. The results of the USPM CFA were χ^2 (44, N = 296) = 330.56, $p < .01$; RMSEA = .15; CFI =. 94; TLI = .92. As with the previous CFA, the χ^2 value was predictably statistically significant due to the sample size so other fit indices were evaluated. These indices, although not as convincing as those from the previous measurement models, indicated the universal screening and progress monitoring latent variable provided a reasonable level of fit for the observed indicators. A detailed examination of the results revealed the USPM latent factor was highly correlated with the DBSP latent factor, which was not surprising because both factors included concepts associated with gathering, analyzing, and using data to inform the instructional planning process. While the values of these indices did not support a strong fit of the data to the hypothesized model, they did indicate the data provided reasonable fit to the hypothesized model because (a) the λ values ranged between .61 and .90 and were all significant at the .01 level and (b) the SMC values for the 10 indicators ranged from .37 to .82. Overall, the CFA results indicated an adequate measurement model for USPM. Accepting the model without revisions provided support for the relationships identified in the full model where universal screening and progress monitoring practices both informed and influenced the MTSS framework. Results of the CFA for the USPM subscale are provided in Table 14 (see Appendix G for the graphic representation).

Table 14

Confirmatory Factor Analysis for Universal Screening and Progress Monitoring

Indicator	λ	SMC
IPS40	.72*	.51
IPS41	.61*	.37
IPS42	.90*	.82
IPS43	.86*	.74
IPS44	.62*	.39
IPS45	.75*	.56
IPS46	.76*	.57
IPS47	.68*	.47
IPS48	.78*	.62
ISP49	.78*	.61
IPS50	.65*	.42

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Factor Analysis for Family, School, and Community Partnerships

As with previous factors, a CFA was conducted on the FSCP factor, which had three observed items. Participants were asked to rate their perceptions of family, school, and community partnering within their individual schools using a 5-point Likert scale. Specifically, the FSCP items read as follows: (a) The staff at my school increases interactions with parents as a student's needs increase; (b) The staff at my school engages families in conversations about student performance data, at least during each parent-teacher conference; and (c) My school helps families understand student performance data for meaningful conversations about student progress. Because all of the degrees of freedom were employed to run the CFA, the χ^2 value for this factor was 0.00 (0, N = 300) = 0.00, $p < .01$ and indicated a perfect fit, which is common for latent variables containing only three indicators (Ulmer, 2004). The λ values were .65, .84, and .99 for each of the items and the SMC values were .42, .71, and .99, respectively. These values were all statistically significant at the .01 level. Although the χ^2 value and zero degrees of freedom meant no fit statistics could be calculated, the λ and SMC values all indicated the fit of the data was sufficient (see Table 15 and Appendix H).

Table 15

Confirmatory Factor Analysis for Family, School, and Community Partnering

Indicator	λ	SMC
IPS37	.65*	.42*
IPS38	.84*	.71*
IPS39	.99*	.99*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Factor Analysis for Layered Continuum of Supports

For LCS, the fifth latent factor of the MTSS framework, a CFA was again conducted to determine the fit of the data. Similar to the previous CFA for the FSCP latent variable, the LCS latent factor consisted of three observed items. Participants were asked to rate their perceptions of the layered continuum of supports component of the MTSS framework using a 5-point Likert scale in their individual schools. The three items for the LCS factor read as follows: (a) The staff at my school regularly meets to determine instructional grouping of students; (b) The curriculum at my school is flexible enough for staff to differentiate instruction based on the individual needs of students; and (c) The staff at my school uses a continuum of increasingly intensive instruction based on student needs and performance levels: all students (universal), some students (targeted), and a few students (intensive). The fit indices for the LCS factor were χ^2 (0, N =315) = 0.00, $p < .01$; RMSEA = .00; CFI = 1.0; TLI = 1.0. As before, the χ^2 value for this factor was 0.00 (0, N = 300) = 0.00, $p < .01$ and indicated a perfect fit, which is common for latent variables containing only three indicators (Ulmer, 2004). The λ values ranged between .62 and .85 and were all significant at the .01 level. The SMC values for the three indicators ranged between .38 and .72. Overall, the CFA results indicated an adequate fit of the data to the measurement model for the LCS. Results of the CFA for the LCS subscale are provided in Table 16 (see Appendix I for the graphic representation).

Table 16

Confirmatory Factor Analysis for Layered Continuum of Supports

Indicator	λ	SMC
IPS27	.73*	.53*
IPS28	.62*	.38*
IPS31	.85*	.72*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation
* $p < .01$.

Factor Analysis for Evidence-Based Practices

For EBP, the sixth latent factor of the MTSS framework, a CFA was again conducted to determine the fit of the data. The EBP scale consisted of five observed items. Participants were asked to rate their perceptions of evidence-based practices component of the MTSS framework using a 5-point Likert scale in their individual schools that primarily focused on best practices for teachers and students alike. Specifically, the items read as follows: (a) My school makes a range of opportunities for coaching and professional development that are aligned to each teacher's specific needs readily available throughout the year; (b) The staff at my school has enough research-based instructional options available to meet the needs of all students; (c) The staff at my school explicitly teaches appropriate behaviors expected of students; (d) When students fail to show appropriate behavior, staff respond by reinforcing the behavioral expectations as they were taught; and (e) The staff at my school engages in classroom management techniques, which creates a positive learning environment for all students.

The CFA of the EBP factor results were $\chi2$ (5, N =315) =85.69, $p < .01$; RMSEA = .23; CFI =. 98; TLI = .97. Once again, the statistically significant $\chi2$ value was not surprising due to the large sample size. While the RMSEA value was higher than expected, the CFI and TLI values indicated the data provided adequate fit to the LCS latent factor. The λ loadings ranged between .59 and .96 and were all statistically significant. The SMC values for the four indicators ranged from .35 to .92 and were statistically significant at the .01 level of significance. Overall, the CFA results indicated an adequate measurement model for the EBP factor. Results of the CFA for the EBP subscale are provided in Table 17 (see Appendix J for the graphic representation).

Table 17

Confirmatory Factor Analysis for Evidence-Based Practices

Indicator	λ	SMC
IPS26	.59*	.35*
IPS30	.62*	.39*
IPS33	.94*	.88*
IPS34	.96*	.92*
IPS35	.85*	.72*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Higher-Order Factor Analyses for the Exogenous Measurement Models

Having determined the data generated by the MTSS-IPS (Pierce et al., n.d.) provided adequate fit to the hypothesized individual latent factors of the MTSS framework, the next phase of the study was to investigate how the individual MTSS factors, both combined and individually, related to third grade reading outcomes. To begin, a variety of models were tested in a series of higher-order CFAs where the data were fit to models with a variety of endogenous latent variables (TDSL, USPM, DBPS, LCS, EBP, FSCP) and MTSS--the single exogenous latent variable. As shared previously, the exogenous variables (i.e., independent variables) were not impacted by any other variable in the model. Conversely, endogenous variables (i.e., dependent variables) are variables that are impacted by exogenous variables. In the discussion that follows, the results of the higher-order CFAs the measurement models with a range of endogenous latent factors and the MTSS systemic framework as the single exogenous latent factor are reported.

Six-Factor Multi-Tiered System of Supports Model

Figure 5 provides a graphic representation of the first higher-order CFA model. Specifically, this model examines the relationships between participants' perceptions of MTSS implementation related to six latent variables of the initial Colorado MTSS model. As a brief review, when the CDE (2016) initially adopted the MTSS framework to provide educators with assistance and support as they worked to maximize the learning outcomes of the students they served, the MTSS model was comprised of six latent factors: (a) team-driven shared leadership (TDSL); (b) evidence-based practices (EBP);

(c) family, school and community partnerships (FSCP); (d) universal screening and progress monitoring (USPM); (e) data-based problem solving (DBPS); and (f) layered continuum of supports (LCS).

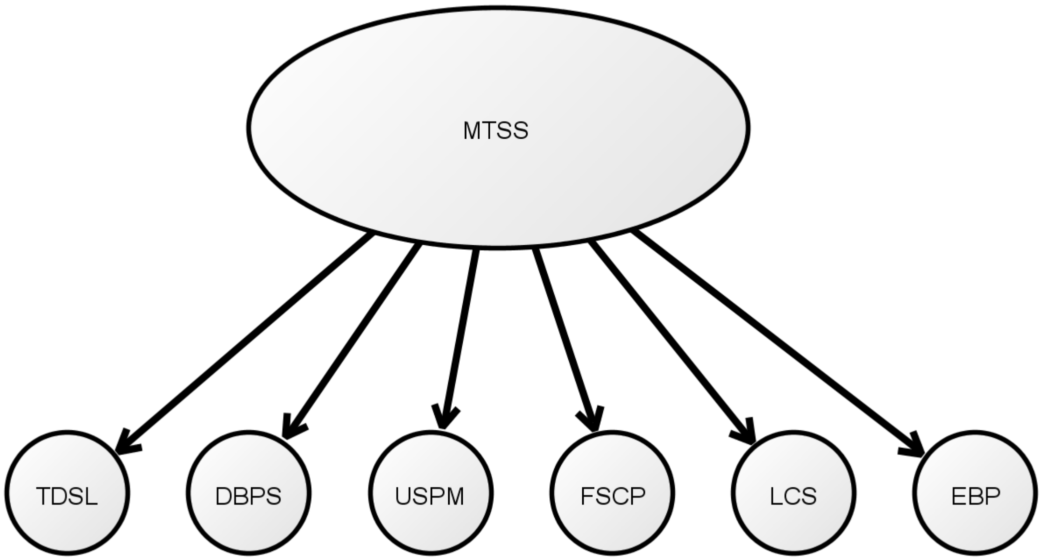


Figure 5. Six-factor Colorado multi-tiered system of supports model.

When assessing the initial fit of the data to this higher-order CFA model, the fit indices demonstrated the six-factor MTSS model provided a good fit for the observed data. Specifically, results for this measurement model were χ^2 (1169, N =382) = 2798.27, p < .01 RMSEA = .06; CFI = .96; TLI = .95. While the χ^2 value was large and statistically significant as shared earlier, previous research has consistently demonstrated this statistic is sensitive to sample size (e.g., Hu & Bentler, 1999). Therefore, other global fit indices were also examined. Recalling the earlier discussion, the recommended cut points for the root mean square error of approximation (RMSEA; Hu & Bentler, 1999) are < .06 for categorical data with smaller values indicating better fit. Conversely, cut-off values for the CFI and the TLI were > .95 and > .96, respectively, with larger values indicating a better fit of the data to the model (Hu & Bentler, 1999). The λ values for the six latent endogenous MTSS factors ranged between .80 and 1.02. All the values fell within the normal range with the exception of the LCS latent factor, which generated undefined results. An investigation of the Mplus results revealed the LCS latent factor was highly correlated with the EBP latent factor. These results indicated that while the data provided adequate fit, specifying an alternative model was appropriate. Results of the CFA for the six-factor MTSS model are provided in Table 18 (see Appendix K for the graphic representation).

Table 18

Higher-Order Factor Analysis for Six-Factor Multi-Tiered System of Supports Framework

Indicator	λ	SMC
TDSL	.93*	.87*
DBPS	96*	.92*
USPM	.84*	.71*
FSCP	.80*	.64*
LCS	1.02*	undefined
EBP	.83*	.69*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* *p* < .01.

Five-Factor Multi-Tiered System of Supports Model

In 2016, leaders within the Office of Learning Supports at the Colorado Department of Education, in an effort to create an effective and efficient MTSS model, subsumed the universal screening and progress monitoring factor within the data-based problem solving latent factor of the MTSS framework. Therefore, the post-2016 CO-MTSS framework is comprised of the following five latent factors: (a) team-driven shared leadership (TDSL); (b) expanded data-based problem solving and decision making (DBPS-e); (c) family, school, and community partnering (FSCP); (d) layered continuum of supports (LCS); and (e) evidence-based practices (EBP; see Figure 6).

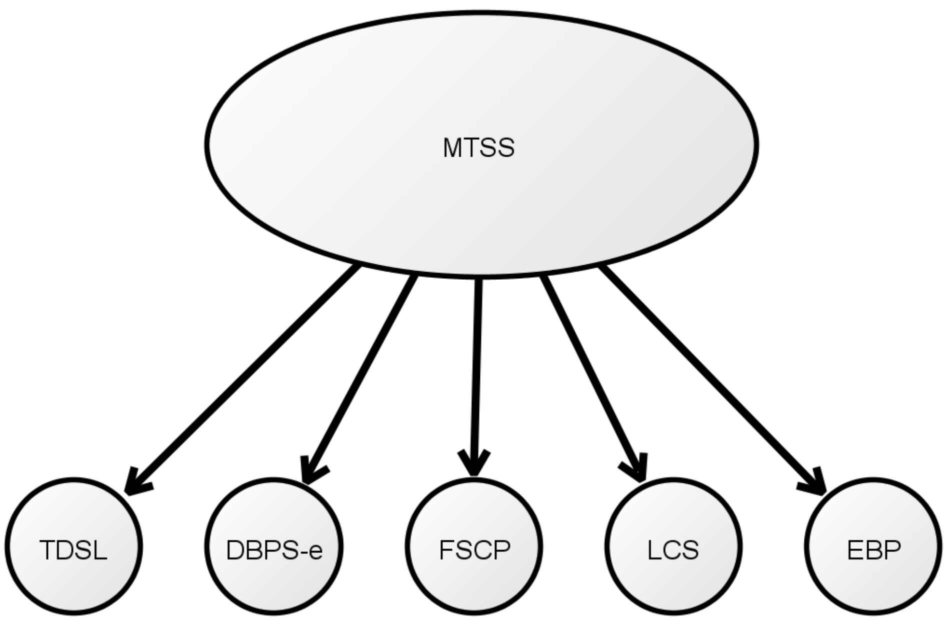


Figure 6. Five-factor multi-tiered system of supports model.

To examine the fit of the MTSS-IPS data to this efficient five-factor MTSS model, indicators that had previously been divided between two factors were merged into a single factor named Expanded Data-based Problem Solving (DBPS-e), which was observed using 25 of the MTSS-IPS items that related participants perceptions' related to both (a) data-based problem solving and decision-making processes and (b) universal screening and progress monitoring implementation efforts within their individual schools (see Table 19).

To examine how the combination of the MTSS-IPS items changed the fit of the data to the expanded DBPS latent factor, a CFA was conducted. The results from this CFA generated the following results: χ^2 (275, N = 377) = 1433.82, p < .01; RMSEA = .11; CFI = .93; TLI = .93. The RMSEA, CFI, and TLI values of .11, .93, and .93, respectively, indicated a poor fit of the data to the extended DBPS latent variable. However, the RMSEA fit statistic was sensitive to the number of parameters estimated in the model so it was unsurprising that its value was above the recommended cut point of .07 recommended by Steiger (2007). However, the λ values for each of the indicator items ranged between .56 and .89 and the SMC values ranged between .31 and .80; these results were similar to previous models that indicated a fair fit of the data to the model. Results from the expanded DBPS latent factor are displayed in Table 20 (see Appendix L for the graphic representation).

Table 19

Expanded Data-Based Problem Solving Item Alignment

IPS No.	Item	Six Factor MTSS-IPS Item Alignment	
		DBPS	USPM
7	The staff at my school uses school-wide achievement trends to decide about interventions and/or instructional strategies for the following year.	✓	
8	The staff at my school analyzes the overall impact of student interventions at the targeted and intensive level at least annually to ensure that the interventions are effective.	✓	
9	My school follows a decision-making process that increases the frequency of progress monitoring as the intensity of instruction and intervention increases.	✓	
16	The staff engaged in problem solving processes at my school works to address the instructional needs of all children in the school, regardless of their academic level.	✓	
17	The staff engaged in problem solving at my school are collectively able to identify appropriate research-based interventions and instructional strategies for students at all academic levels.	✓	
18	The problem solving process at my school allows the staff to adjust instructional supports based on student data/results.	✓	
19	The staff engaged in problem solving at my school uses data to identify individual student need for targeted and intensive intervention.	✓	
20	The staff engaged in problem solving at my school uses data sources in addition to summative data from the state to analyze achievement trends collectively for all students.	✓	
21	The staff at my school use data to evaluate the effectiveness of our math curriculum.	✓	
23	The staff at my school use data to evaluate and improve their own instructional practices.	✓	
24	The staff at my school works collaboratively to use data to assess and support their peers for continuous improvement of instructional practices.	✓	
25	The staff at my school collects and analyzes information to determine whether differentiation of instruction occurs based on student need.	✓	
29	Defined decision-making processes at my school enable the staff to efficiently select interventions or instruction based on the level of student need.	✓	
32	Members of my schools problem solving team have clear roles and responsibilities.	✓	
40	My school has a data management system for tracking academic progress of all students that is functional, useful, and accessible by all staff		✓
41	My school has a data management system to track school-wide behavior data (e.g., discipline referrals, truancy, attendance) that is functional, useful, and accessible by all staff.		✓
42	The staff at my school is proficient in accessing achievement data for our students.		✓
43	The staff at my school knows how to interpret data to inform instructional practices.		✓
44	The staff at my school uses standardized formative assessments (e.g., AIMSweb, Galileo, NWEA) to monitor student progress.		✓

Table 19 Continued

IPS No.	Item	Six Factor MTSS-IPS Item Alignment	
		DBPS	USPM
45	The staff at my school uses informal classroom formative assessments (e.g., observations, classroom quizzes, exit tickets, walk-arounds) to identify the immediate instructional needs of our students.		✓
46	The staff at my school uses universal screening measures to identify any students needing additional supports to progress from their current academic level (e.g., accelerated, delayed, etc.).		✓
47	My school administers universal screening and benchmarking assessments in math at regular intervals.		✓
48	My school's assessment system provides guidelines on types of data needed to establish a body of evidence for eligibility for gifted services.		✓
49	My school’s assessment system provides guidelines on types of data needed to establish a body of evidence for eligibility for all categories of special education.		✓
50	All students at my school are involved in monitoring their own progress for the purpose of setting their own academic goals.		✓

Note. DBPS= Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring.

Table 20

Confirmatory Factor Analysis for Expanded Data-Based Problem Solving Factor

Indicator	λ	SMC
IPS7	.79*	.63*
IPS8	.84*	.71*
IPS9	.83*	.69*
IPS16	.86*	.74*
IPS17	.88*	.77*
IPS18	.88*	.78*
IPS19	.89*	.80*
IPS20	.78*	.61*
IPS21	.72*	.52*
IPS23	.82*	.68*
IPS24	.84*	.70*
IPS25	.82*	.67*
IPS29	.85*	.72*
IPS32	.77*	.60*
IPS40	.65*	.42*
IPS41	.57*	.33*
IPS42	.82*	.67*
IPS43	.84*	.70*
IPS44	.56*	.31*
IPS45	.72*	.52*
IPS46	.69*	.48*
IPS47	.56*	.32*
IPS48	.70*	.49*
IPS49	.69*	.48*
IPS50	.61*	.37*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Next, a higher-order CFA that investigated the adequacy of the fit of the data to the more efficient five-factor MTSS model was conducted and generated the following results: χ^2 (1170, N =382) = 3164.89, $p < .01$; RMSEA = .07; CFI = .95; TLI = .94, which indicate the MTSS-IPS data provided a better fit to the six-factor MTSS model than the efficient five-factor MTSS model. All the values fell within the normal range with the exception of the LCS latent factor. As with the previous model, the parameter estimates were again undefined for the LCS variable. Results from the five-factor MTSS CFA are displayed in Table 21 (see Appendix M for the graphic representation).

Table 21

Confirmatory Factor Analysis for Five-Factor Multi-Tiered System of Supports Framework

Indicator	λ	SMC
TDSL	.93*	.87*
DBPS-e	.96*	.92*
FCSP	.80*	.64*
LCS	1.02*	Undefined
EBP	.83*	.69*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Revised Five-Factor Multi-Tiered System of Supports Model

Previous results revealed the LCS latent variable consistently produced out of range parameter estimates in both the five- and six-factor models. An examination of the data output revealed the LCS highly correlated with a variety of the other latent factors included in the model (e.g., TDSL, DBPS, EBP), generating excessive R^2 values of .99. Therefore, the LCS latent factor was removed from the previous higher-order six-factor CFA (see Figure 7).

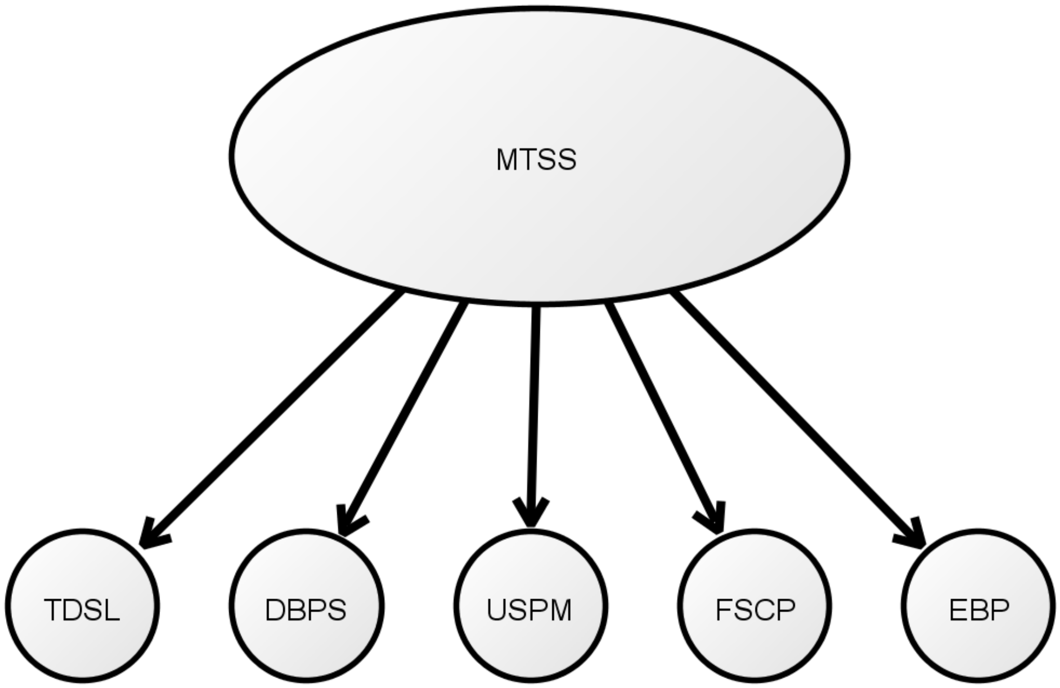


Figure 7. Revised five-factor multi-tiered system of supports model.
Note. This revised five-factor model is the result of removing the highly correlated LCS latent factor from the previous six-factor MTSS model. TDSL = Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP = Family, School, and Community Partnerships; EBP = Evidence-Based Practices.

Results of the higher-order CFA for the revised five-factor model depicted in Figure 7 indicated the revised model provided a better fit for the observed data than the previous five-factor model that included the LCS latent factor, specifically, χ^2 (1029, N =382) = 2484.58, p < .01; RMSEA = .06; CFI = .96; TLI = .95. The λ values for the revised five-factor model ranged between .81 and .95 and the SMC values ranged between .65 and .91--all of which fell well within in the acceptable range of estimates. Results of the revised model are provided in Table 22 and a graphic representation of the model results is provided in Appendix N.

Table 22

Confirmatory Factor Analysis for Revised Five-Factor Multi-Tiered System of Supports Framework

Indicator	λ	SMC
TDSL	.94*	.88*
DBPS	.95*	.91*
USPM	.84*	.71*
FCSP	.81*	.66*
EBP	.83*	.69*

Note. λ= Completely Standardized Factor Loading; SMC = Squared Multiple Correlation.
* $p < .01$.

Four-Factor Multi-Tiered System of Supports Model

To investigate how removing the LCS latent factor impacted the fit of the data to the previous five-factor model that included the expanded DBPS latent factor, a four-factor model was identified and a fourth higher-order CFA was conducted (see Figure 8).

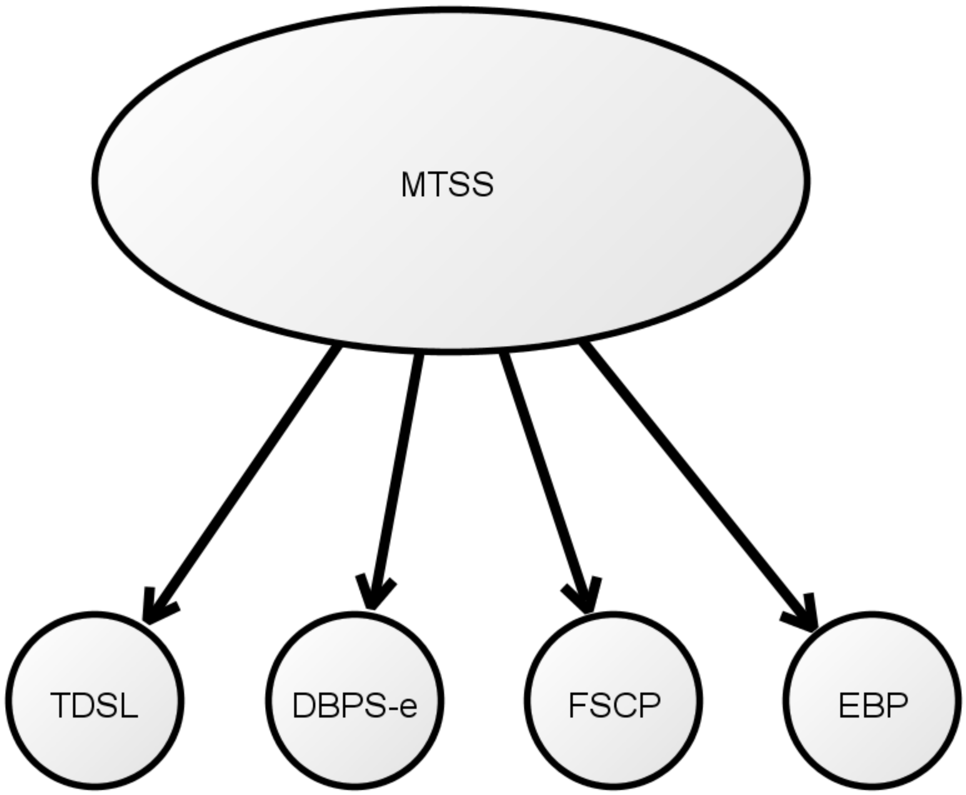


Figure 8. Four-factor multi-tiered system of supports model.
Note. This revised four-factor model is the result of removing the highly correlated LCS latent factor from the previous five-factor MTSS model represented in Figure B. TDSL = Team-Driven Shared Leadership; DBPS-e = Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP = Evidence-Based Practices.

When the fit of the observed data to the revised four-factor MTSS model was evaluated, the results indicated the revised five-factor model reported previously provided a better fit for the MTSS-IPS data. Specifically, results of the four-factor model were χ^2 (1030, N =382) = 2962.69, p < .01; RMSEA = .07; CFI = .94; TLI = .94, which all indicated the model provided a worse fit for the data than previous models. The λ values for the revised four-factor model were identical to the revised five-factor higher order CFA, ranging between .81 and .95. Similarly, the SMC values ranged between .65

and .91. Results of the fourth higher-order CFA are provided in Table 23 and Appendix O.

Table 23

Confirmatory Factor Analysis for Four-Factor Multi-Tiered System of Supports Framework

Indicator	λ	SMC
TDSL	.94*	.88*
DBPS-e	95*	.91*
FCSP	.81*	.66*
EBP	.83*	.69*

Note. λ= Completely Standardized Factor Loading; SMC= Squared Multiple Correlation.
* $p < .01$

Comparison of the Higher-Order Factor Analyses

In an examination of the higher-order CFA that examined the fit of the MTSS-IPS data to the theorized models, the results indicated the revised five-factor model, which separated the data-based problem solving items from the items conceptually linked with the universal screening and progress monitoring, provided a closer fit for the data. Table 24 provides a summary of the results by model. In the two models that included the highly correlated LCS latent factor, model fit tended to improve when additional latent factors were included. Similarly, when the fit of the data was compared to the two models that did not include the LCS latent factor, fit of the data improved in the model that included more latent variables and used more of the available degrees of freedom.

Finally, a comparison of the χ^2 values between the six- and revised-five factor models favored the revised five-factor model because smaller χ^2 values resulted, indicating a better fitting model (Suhr, 2008).

Table 24

Comparison of Higher-Order Confirmatory Factor Analyses

Model	LV	*df*	χ^2	RMSEA	CFI	TLI
Six-Factor	TDSL, DBPS, USPM, EBP, FSCP, LCS	1169	2798.27*	.06	.96	.95
Five-Factor	TDSL, DBPS-e, EBP, FSCP, LCS	1170	3164.89*	.07	.95	.94
Five-Factor (revised)	TDSL, DBPS, USPM, EBP, FSCP	1029	2584.58*	.06	.96	.95
Four-Factor	TDSL, DBPS-e, EBP, FSCP	1030	2962.69*	.07	.94	.94

Note. Six-factor = Pre-2016 Colorado Six-factor MTSS Model; Five-Factor = Post-2016 Colorado Five-Factor MTSS model; Five-Factor Revised = Six-Factor model without the LCS factor; Four-Factor = Five-Factor without the LCS factor; LV= Individual latent variables included in the model; df= Degrees of Freedom; χ^2= Chi-square test of model fit; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index; TLI= Tucker Lewis Index.
* $p < .01$

Higher-Order Structural Models

As noted earlier, the primary purpose of this study was to investigate the relationship between third grade reading achievement and MTSS implementation perceptions in the state of Colorado. Having determined the MTSS-IPS (Pierce et al., n.d.) generated data that were valid, reliable, insensitive to non-response bias, and contributed in a meaningful way to the proposed MTSS frameworks, a series of higher-order SEMs were conducted to examine the relationship between the latent factors of the

MTSS models and 2014 Colorado third grade reading outcomes. Publicly available 2014 third grade TCAP reading data were used as the endogenous (dependent) observed variable of primary interest while MTSS was the exogenous (independent) latent variable in all of the structural models. The results of these analyses are presented in the following sections.

Six-Factor Higher-Order Structural Model

To examine the fit of the data to the higher-order full structural model that included school-level 2014 third grade TCAP reading scores as the endogenous variable of interest, the initial fit of the data to the original six-factor higher-order MTSS framework was conducted (see Figure 9).

The global fit indices demonstrated the data provided a good fit to the six-factor model that included a second-order factor model with TCAP scores as the endogenous variable in the model even though the highly-correlated LCS latent variable was included. Specifically, results for this measurement model were χ^2 (1218, N = 511) = 2856.51, p < .01; RMSEA = .05; CFI = .96; TLI = .95. The RMSEA fit index value of .05 indicated good fit and the values for both the CFI and TLI indices met minimum criteria (Schreiber et al., 2006). Notably, the completely standardized path coefficient, β, that measured the direct effect of MTSS on 2014 third grade TCAP reading scores was positive and statistically significant (β = .18, SE = .06). This seemed to indicate perceptions of MTSS were positively related to 2014 third grade TCAP reading scores. The ranges for each standardized factor loading that quantified the indicator by latent factor relationship, the model fit indices, and the path coefficient between MTSS and TCAP scores are included in Table 25 (see Appendix P for the graphic representation).

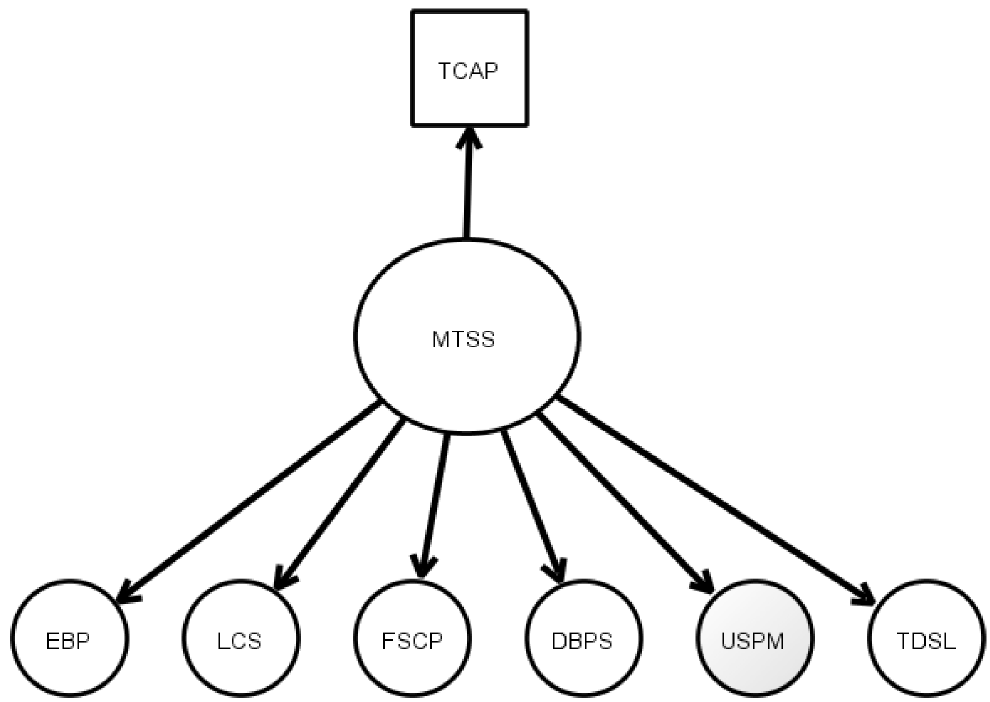


Figure 9. Six-factor higher-order structural model and transitional Colorado academic program.

Note. TCAP = 2014 third grade TCAP reading scores; MTSS = Pre-2016 Colorado Multi-tiered System of Supports framework; EBP= Evidence-Based Practice; LCS = Layered Continuum of Supports; FSCP= Family, School, and Community Partnerships; DBPS= Data-Based Problem Solving; USPM= Universal Screening and Progress Monitoring; TDSL= Team-Driven Shared Leadership.

Table 25

Six-Factor Higher-Order Structural Model

LV	IPS Items	λ (Range)	SMC(Range)
TDSL	1-6, 10-15, 22, 36	.74 - .92*	.54 - .85*
DBPS	7-9, 16-21, 23-25, 29, 32	.72 - .89*	.52 - .79*
USPM	40-50	.59 - .92*	.35 - .84*
FSCP	37-39	.83 - .91*	.68 - .82*
LCS	27, 28, 31	.63 - .85*	.39 - .72*
EBP	26, 30, 33-35	.83 - .95*	.68 - .90*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$;

Five-Factor Higher-Order Structural Model

After the fit of the data to the higher-order six factor structural model was evaluated, a second higher-order SEM was conducted to assess the fit of the data to a higher-order five-factor MTSS model, which once again included the highly correlated LCS latent factor to examine if the more condensed model's results provided a better fit for the subset of Colorado-specific data generated by MTSS-IPS (see Figure 10).

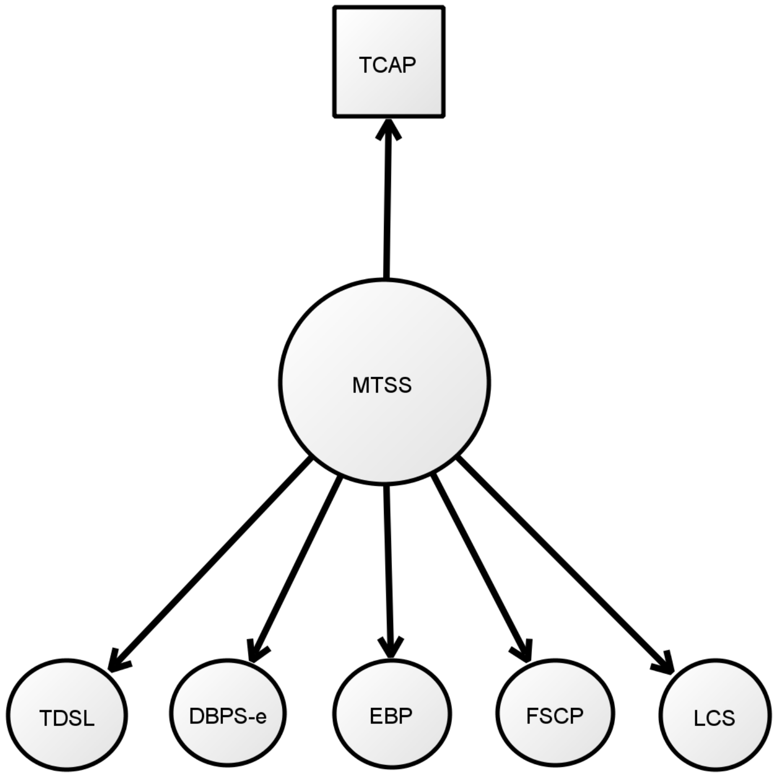


Figure 10. Five-factor higher-order structural model and transitional Colorado academic program.
Note. TCAP = 2014 third grade TCAP reading scores; MTSS = Current Colorado Multi-tiered System of Supports framework; TDSL= Team-Driven Shared Leadership; DBPS-e= Expanded Data-Based Problem Solving; EBP= Evidence-Based Practice; FSCP= Family, School, and Community Partnerships; LCS = Layered Continuum of Supports.

The results of that analysis were χ^2 (1219, N = 511) = 3218.72, p <.01; RMSEA = .06; CFI = .95; TLI = .94. As in the previous six-factor higher-order SEM, the standardized path coefficient between MTSS and TCAP was small but positive and statistically significant (β = .18), which indicated that as perceptions of MTSS increased, third grade 2014 TCAP scores also tended to increase. An inspection of the λ values indicated the range of factor loading for four of the five LV remained relatively stable. The largest change in factor loadings was for the expanded DBPS factor, which included previous items forced to load on the USPM latent factor. Similar results were obtained for the SMC for each of the items. Results are provided in Table 26 (see Appendix Q for the graphic representation)

Table 26

Five-Factor Higher-Order Structural Model

LV	IPS Items	λ (Range)	SMC(Range)
TDSL	1-6, 10-15, 22, 36	.74 - .92*	.55 - .85*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53 - .88*	.28 - .77*
FSCP	37-39	.83 - .91*	.69 - .83*
LCS	27, 28, 31	.63 - .85*	.40 - .72*
EBP	26, 30, 33-35	.83 - .95*	.69 - .90*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Revised Five-Factor Higher-Order Structural Model

To investigate the fit of the MTSS-IPS data to the revised five-factor MTSS higher-order structural model (see Figure 11), a third set of analyses was conducted.

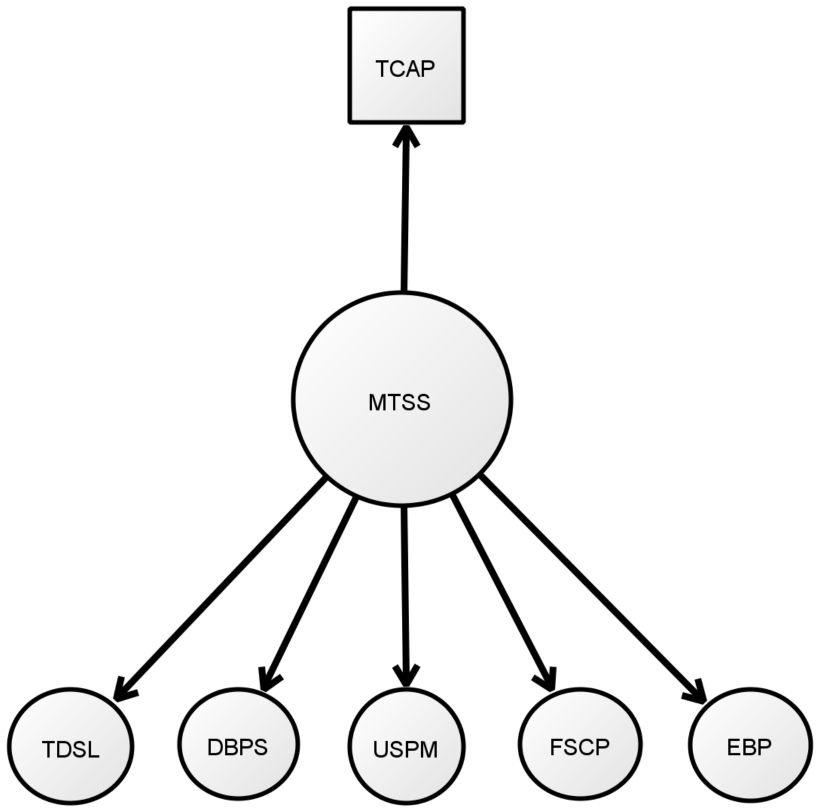


Figure 11. Revised five-factor higher-order structural model and transitional Colorado academic program.
Note. TCAP = 2014 third grade TCAP reading scores; MTSS = Revised Five-Factor Colorado Multi-tiered System of Supports Framework; TDSL= Team-Driven Shared Leadership; DBPS= Data-Based Problem Solving; USPM= Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

This revised five-factor higher-order SEM, which did not include the highly correlated LCS latent endogenous factor, provided a good level of fit for the data; it was similar to the fit provided by the six-factor higher-order model (see Table 25). Specifically, the results were χ^2 (1075, N = 511) = 2640.68, p < .01; RMSEA = .05; CFI = .96; TLI = .96. When compared with the previous model, the factor loadings of the individuals differed from the original five-factor model in the data but were similar to the factor loadings of the six-factor model. Similarly, the standardized path coefficient between MTSS and TCAP was small but positive, statistically significant, and indicated perceptions of MTSS implementation appeared to be positively related to the 2014 third grade TCAP reading scores of the schools that were included in the sample ($\beta = .18$). Results of the revised five-factor MTSS SEM are provided in Table 27 (see Appendix R for the graphic representation).

Table 27

Revised Five-Factor Higher-Order Structural Model

LV	IPS Items	λ (Range)	SMC(Range)
TDSL	1-6, 10-15, 22, 36	.74 - .92*	.55 - .85*
DBPS	7-9, 16-21, 23-25, 29, 32	.73 - .89*	.28 - .77*
USPM	40-50	.59 - .91*	.34 - .84*
FSCP	37-39	.83 - .91*	.68 - .82*
EBP	26, 30, 33-35	.83 - .93*	.69 - .87*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations
* $p < .01$.

Four-Factor Higher-Order Structural Model

The final higher-order structural model that simply examined the fit of the data to a model that contained the expanded DBPS factor, but did not include the LCS factor provided an adequate fit for the data but did not exceed the level of fit for the revised five-factor or original six-factor higher-order SEM. Figure 12 provides a graphic representation of the four-factor model.

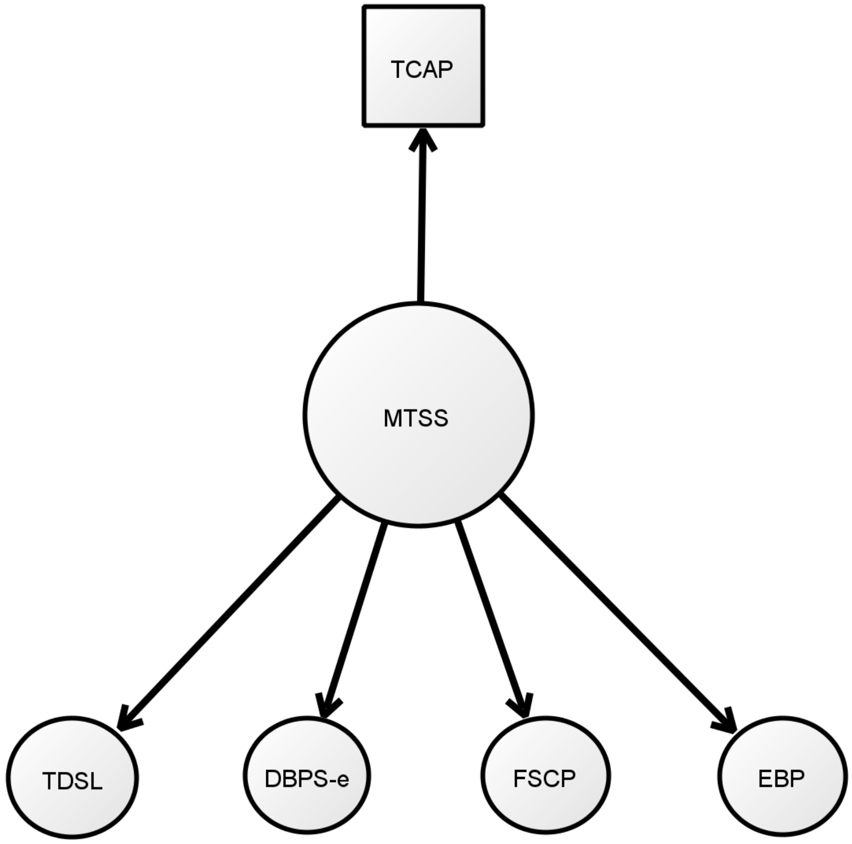


Figure 12. Four-factor higher-order structural model and transitional Colorado academic program.
Note. TCAP = 2014 third grade TCAP reading scores; MTSS = Current Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-driven shared leadership; DBPS-e= Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

The fit of the data to the four-factor higher order structural model revealed the model did not provide a better fit than the revised five-factor model; χ^2 (1076, N = 511) = 3013.87, p <.01; RMSEA = .06; CFI = .95; TLI = .94. As in all the previous higher-order SEMs, the standardized path coefficient between TCAP and MTSS was positive and statistically significant (β = .18). Results of the analysis are summarized in Table 28 (see Appendix S for the graphic representation).

Table 28

Four-Factor Higher-Order Structural Model

LV	IPS Items	λ(Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .89*	.55 - .86*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53 - .88*	.28 - .78*
FSCP	37-39	.83 - .91*	.68 - .82*
EBP	26, 30, 33-35	.83 - .93*	.68 - .87*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$

Comparison of the Higher-Order Structural Models

The second research question this study posed asked if one of the higher-order SEM theorizing relationships between perceptions of MTSS implementation in Colorado and 2014 Colorado third grade reading achievement provided a better fit for the observed data. In an examination of the higher-order SEMs that analyzed the fit of the MTSS-IPS data to the theorized higher-order structural models, the results indicated the revised five-factor model that separated the data-based problem solving items from the items conceptually linked with universal screening and progress monitoring provided a closer fit for the data. In the two models that included the highly correlated LCS latent factor, model fit tended to improve when additional latent factors were included. Similarly, when the fit of the data was examined to the two models that did not include the LCS latent factor, the model that included more latent variables provided a better fit. In all the higher-order SEMs, the path coefficients between TCAP and MTSS remained stable, statistically significant, and positive. Results of the four higher-order structural models are summarized in Table 29.

Table 29

Comparison of Higher-Order Structural Models

SEM	LV	*df*	χ^2	RMSEA	CFI	TLI
Six-factor	TDSL, DBPS, USPM, EBP, FSCP, LCS	1218	2856.51*	.05	.96	.95
Five-factor	TDSL, DBPS-e, EBP, FSCP, LCS	1219	3218.72*	.06	.95	.94
Five-factor (revised)	TDSL, DBPS, USPM, EBP, FSCP	1075	2640.68*	.05	.96	.96
Four-factor	TDSL, DBPS-e, EBP, FSCP	1076	3013.87*	.06	.95	.94

Note. Six-factor = Pre-2016 Colorado Six-factor MTSS Model; Five-factor = Post-2016 Colorado Five-factor MTSS model; Five-factor Revised = Six-factor without the LCS factor; Four-factor = Five-factor without the LCS factor; LV= Individual latent variables included in the model; *df*= Degrees of Freedom; χ^2= Chi-square test of model fit; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index; TLI= Tucker Lewis Index.
* $p < .01$

The analysis above provided evidence the revised five-factor higher order hypothesized MTSS model provided the best fit for the MTSS-IPS data. Specifically, RMSEA values for the revised five-factor model were .05 and the CFI values were .96, which were identical to those generated for the six-factor model. However, the smaller value of the χ^2 statistic (2640.68) and the normed TLI fit index (.96) indicated the revised five-factor model provided a better fit for the data than any of the other proposed models.

Higher-Order Structural Models with Free and Reduced Lunch

To investigate the effect that including an indicator of SES at the school-level had on the perception of MTSS implementation and 2014 third grade TCAP reading outcomes, the percentage of students who qualified for free and reduced lunch (FRL) within each of the schools included in the sample was added as an exogenous variable in a second set of higher-order SEM models. Results from each of the second set of higher-order structural models are discussed in the following sections.

Six-Factor Higher-Order Structural Model with Free and Reduced Lunch

When assessing the initial fit of the data to the six-factor higher-order MTSS framework that included FRL as an exogenous variable using SEM techniques (see Figure 13), the global fit indices demonstrated the data provided good fit to the hypothesized MTSS higher-order model even when the highly correlated LCS latent variable was included in the model.

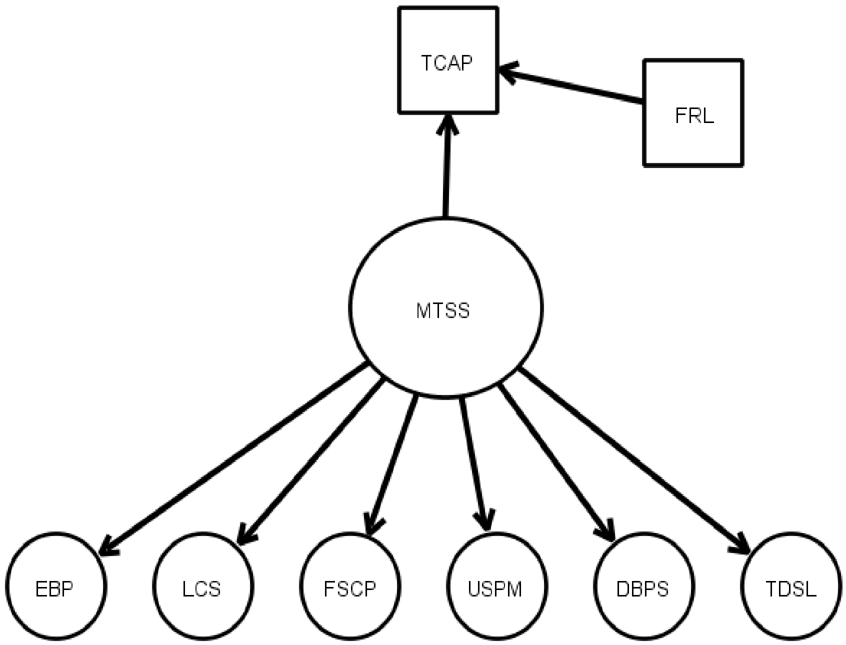


Figure 13. Six-factor higher-order structural model with free and reduced lunch.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; MTSS = Pre-2016 Colorado Six-factor Multi-tiered System of Supports Model; EBP= Evidence-Based Practice; LCS = Layered Continuum of Supports; FSCP= Family, School, and Community Partnerships; USPM = Universal Screening and Progress Monitoring; DBPS = Data-Based Problem Solving; TDSL= Team-Driven Shared Leadership.

Results for this measurement model were χ^2 = 2709.98 (1268, N =508), p < .01; RMSEA = .05; CFI = .96; TLI = .96. The standardized path coefficient between TCAP and FRL (β = -.67) was statistically significant and negative, which indicated the 2014 third grade TCAP reading scores tended to decrease as the percentage of students who qualified for free and reduced lunch increased. However, even with the addition of the FRL exogenous variable, the path between TCAP and MTSS while small remained statistically significant (β = .12). This seemed to indicate as perceptions of MTSS implementation increased, school-level 2014 third grade TCAP reading scores also increased. An examination of the λ values for each observed item and the individual factors revealed each item had a high loading factor value. The ranges for each standardized indicator by latent factor as well as ranges for SMC values are included in Table 30 (see Appendix T for the graphic representation).

Table 30

Six-Factor Higher-Order Structural Model with Free and Reduced Lunch

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .93*	.55 - .86*
DBPS	7-9, 16-21, 23-25, 29, 32	.72 - .89*	.52 - .79*
USPM	40-50	.59 - .92*	.35 - .85*
FSCP	37-39	.83 - .91*	.69 - .82*
LCS	27, 28, 31	.64 - .85*	.41 - .72*
EBP	26, 30, 33-35	.83 - .96*	.69 - .92*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Five-Factor Higher-Order Structural Model with Free and Reduced Lunch

Next, the initial fit of the data to the original five-factor higher-order MTSS framework that included FRL as an exogenous variable was conducted (see Figure 14).

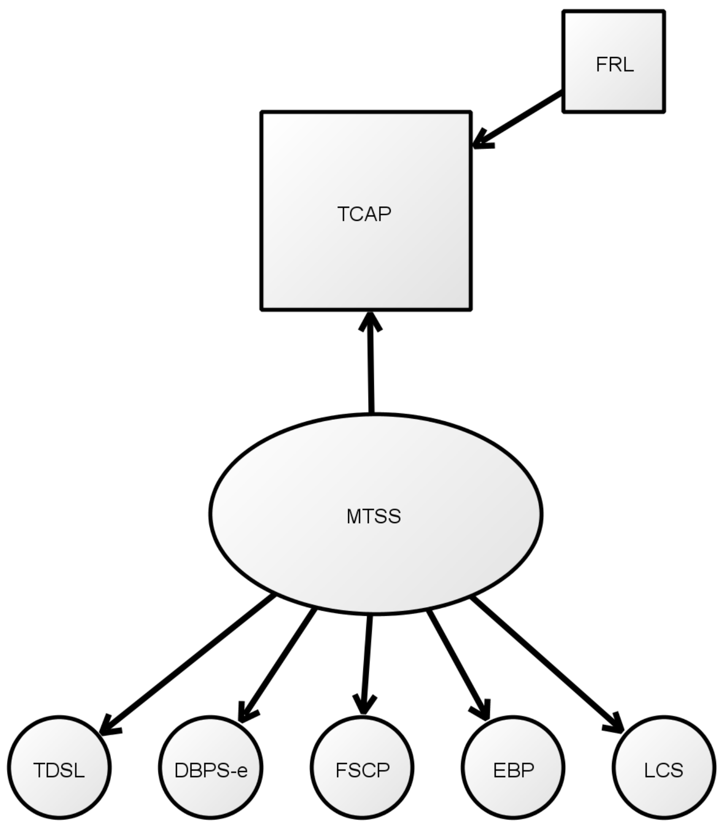


Figure 14. Five-factor higher-order structural model with free and reduced lunch. *Note.* TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; MTSS = Current Colorado Multi-tiered System of Supports framework; TDSL= Team-Driven Shared Leadership; DBPS-e=Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice; LCS = Layered Continuum of Supports.

As in the previous original five-factor models, this model included the expanded DBPS latent factor as well as the troublesome, highly correlated, and non-positive definite LCS latent endogenous factor. The global fit indices generated from this analysis demonstrated that again the data provided good fit to the model. Specifically, results for this five-factor measurement model were χ^2 (1269, *N* =508) = 3053.32, $p < .01$; RMSEA = .05; CFI = .95; TLI = .95. Similar to the previous six-factor higher-order SEM that included FRL as an exogenous variable, the standardized path coefficient between FRL and MTSS was negative and statistically significant ($\beta = -.67$), while the path between MTSS and TCAP was positive and statistically significant ($\beta = .12$). An examination of the λ values for each observed item and the individual factors demonstrated that each item has a high loading factor value. The ranges for each standardized indicator by latent factor, as well as ranges for SMC values are included in Table 31 (see Appendix U for the graphic representation).

Table 31

Five-Factor Higher-Order Structural Model with Free and Reduced Lunch

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.76 - .93*	.58 - .86*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53 - .89*	.28 - .79*
FSCP	37-39	.83 - .91*	.69 - .83*
LCS	27, 28, 31	.64 - .85*	.41 - .72*
EBP	26, 30, 33-35	.83 - .96*	.68 - .92*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Revised Five-Factor Higher-Order Structural Model with Free and Reduced Lunch

Next, the fit of the data to the revised five-factor model that included the USPM and DBPS factors but eliminated the LCS factor was assessed (see Figure 15). Similar to the previous revised five-factor higher-order models with and without exogenous variables, this revised five-factor model also provided a good fit for the data. Specifically, the statistics and fit indices were χ^2 (1122, N =508) = 2477.72, p < .01; RMSEA = .05; CFI = .96; TLI = .96. As in the previous structural models that included FRL as an exogenous variable, the path between MTSS and FRL was large, negative, and statistically significant (β = -.67), while the path between MTSS and TCAP remained small but statistically significant (β = .12). Results are summarized in Table 32 and displayed in Appendix V.

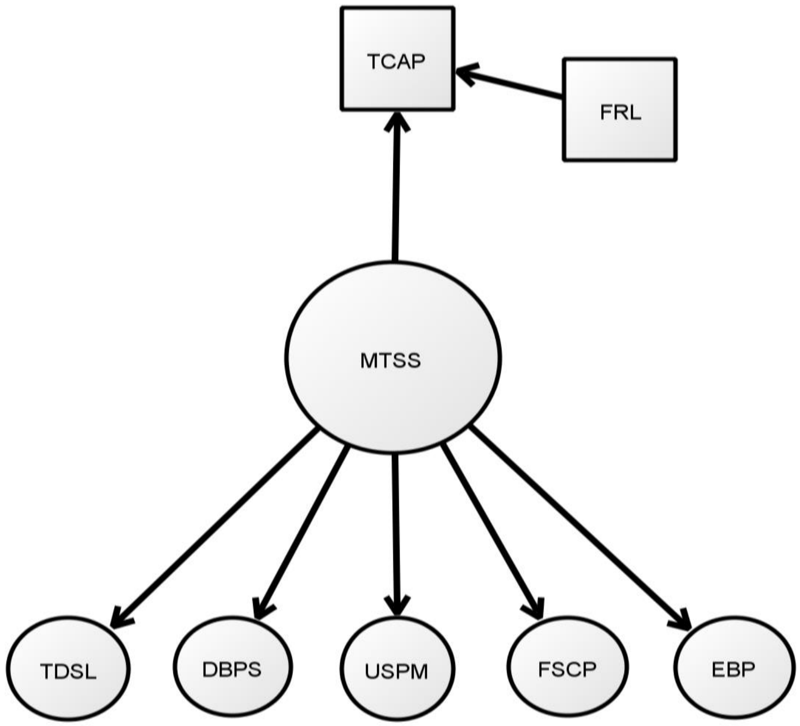


Figure 15. Revised five-factor higher-order structural model with free and reduced lunch.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; MTSS = Pre-2016 Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM= Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

Table 32

Revised Five-Factor Higher-Order Structural Model with Free and Reduced Lunch

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74-.93*	.53-.86*
DBPS	7-9, 16-21, 23-25, 29, 32	.73-.90*	.53-.81*
USPM	40-50	.59-.92*	.35-.85*
FSCP	37-39	.83-.91*	.69-.83*
EBP	26, 30, 33-35	.83-.94*	.68-.88*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Four-Factor Higher-Order Model with Free and Reduced Lunch

The final higher-order structural model that simply examined the fit of the data to a model that contained the expanded DBPS factor but did not include the LCS factor provided an adequate fit for the data but did not exceed the level of fit for the revised five-factor or the original six-factor higher-order SEMs (see Figure 16).

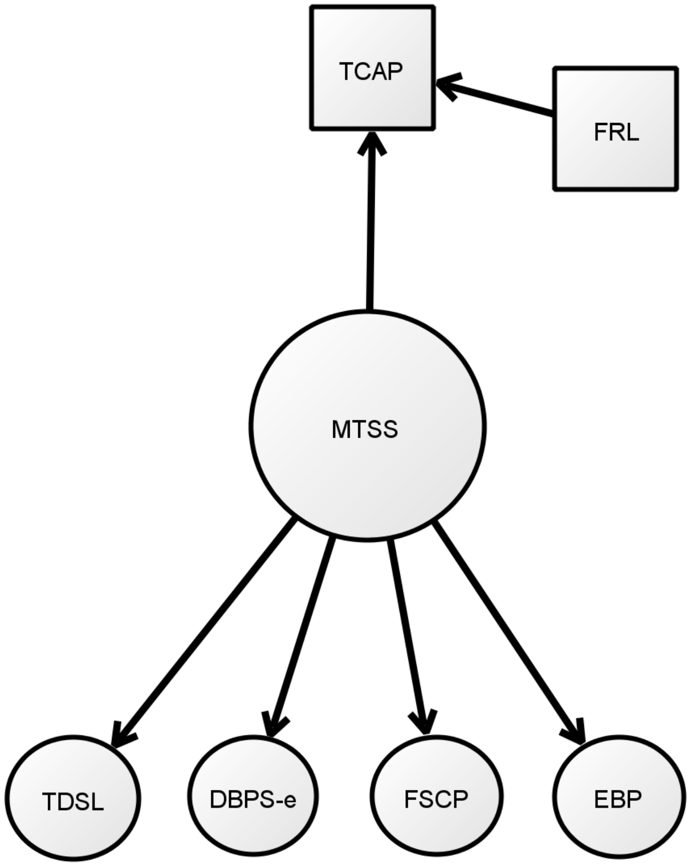


Figure 16. Four-factor higher-order structural model with free and reduced lunch.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; MTSS = Current Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-driven shared leadership; DBPS-e = Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

As before, the fit the data to a final four-factor model that included the expanded DBPS factor, excluded the LCS factor, and included the FRL as an exogenous indicator was examined. Results indicated the model did not provide a better fit for the data than the revised five-factor mode; χ^2 (1123, N = 508) = 2828.97, p <.01; RMSEA = .06; CFI = .95; TLI = .95. Similar to the previous structural models that included FRL as an exogenous variable, the standardized path coefficient from TCAP and FRL was large, negative, and statistically significant (β = -.67), while the standardized path coefficient between MTSS and TCAP remained small but statistically significant (β = .12). Results of the analysis are summarized in Table 33 and displayed in Appendix W.

Table 33

Four-Factor Higher-Order Structural Model with Free and Reduced Lunch

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .93*	.55 - .87*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53 - .89*	.28 - .79*
FSCP	37-39	.83 - .91*	.68 - .82*
EBP	26, 30, 33-35	.83 - .94*	.68 - .88*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Comparison of the Higher-Order Structural Models with Free and Reduced Lunch

Having analyzed the series of higher-order SEMs to examine the fit of the MTSS-IPS to 2014 third grade TCAP data when FRL was included as an exogenous variable, a comparative analysis of the fit of the data to the models was again possible. Similar to the earlier comparison, the statistics indicated the revised five-factor model provided a better fit for the MTSS-IPS data. Specifically, RMSEA values for the revised five-factor model (.05) and the CFI values (.96) were identical to those generated for the six-factor model. However, the smaller value of the $\chi 2$ statistic and the normed TLI fit index indicated the revised five-factor model provided a better fit for the data than any of the other proposed models. Results of the four higher-order SEMs with FRL are summarized in Table 34.

Table 34

Comparison of Higher-Order Structural Models with Free and Reduced Lunch

SEM	LV	*df*	χ^2	RMSEA	CFI	TLI
Six-factor	TDSL, DBPS, USPM, EBP, FSCP, LCS	1268	2709.98*	.05	.96	.96
Five-factor	TDSL, DBPS-e, EBP, FSCP, LCS	1269	3053.32*	.05	.95	.95
Five-factor (revised)	TDSL, DBPS, USPM, EBP, FSCP	1122	2477.72*	.05	.96	.96
Four-factor	TDSL, DBPS-e, EBP, FSCP	1123	2828.97*	.06	.95	.95

Note. Six-factor = Pre-2016 Colorado Six-factor MTSS Model; Five-factor = Post 2016 Colorado Five-factor MTSS model; Five-factor Revised = Six-factor without the LCS factor; Four-factor = Five-factor without the LCS factor; LV= Individual latent variables included in the model; *df*= Degrees of Freedom; χ^2= Chi-square test of model fit; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index; TLI= Tucker Lewis Index.
* $p < .01$.

Higher-Order Structural Models with District Size

To investigate how much of the variance in 2014 third grade TCAP reading scores could be accounted for by district-level funding, the district size (DS) of the schools included in the sample was included as an exogenous observed variable. Because the student enrollment of the districts included in the sample ranged between 12 and 86,043, this variable was standardized. Results from each of the second set of higher-order structural models are discussed in the following sections.

Six-Factor Higher-Order Structural Model with District Size

When assessing the initial fit of the data to this six-factor higher-order MTSS framework that included DS as an exogenous variable using SEM techniques, the global fit indices demonstrated the data provided good fit to the hypothesized MTSS higher-order model (see Figure 17) even when the highly correlated LCS latent variable was included in the model.

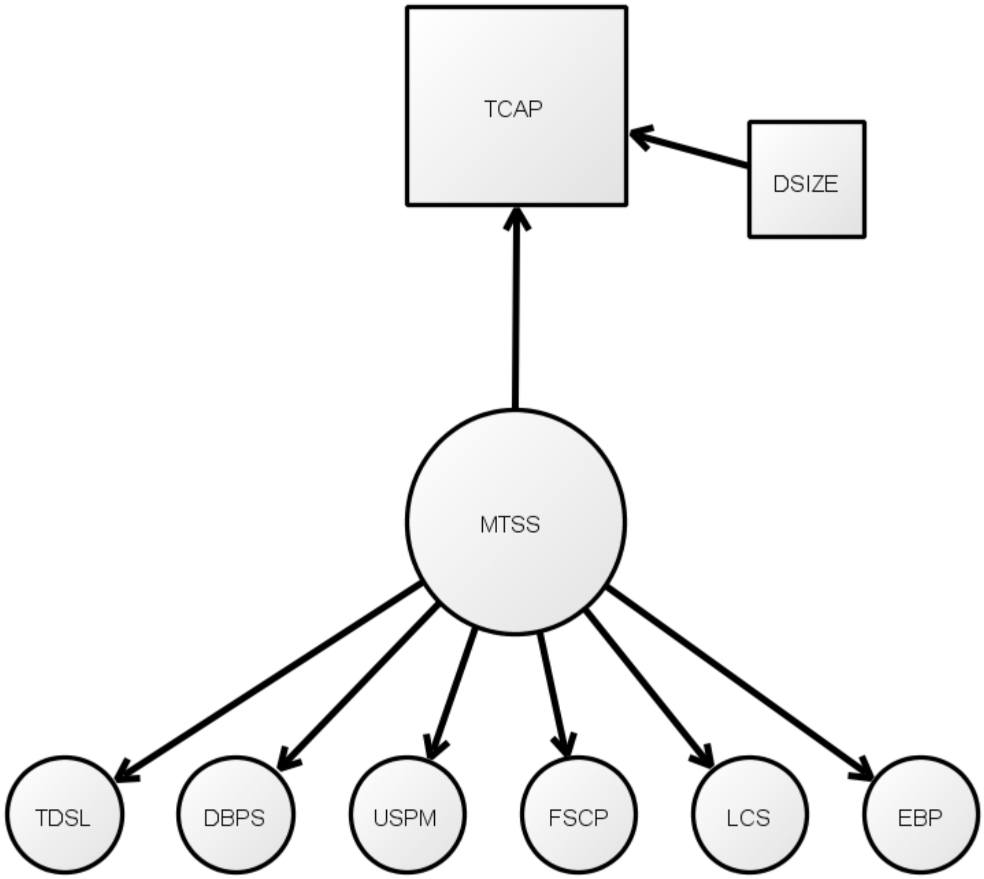


Figure 17. Six-factor higher-order structural model with district size.
Note. TCAP = 2014 third grade TCAP reading scores; DSize= Standardized K-12 enrollment for each school's district; MTSS = Pre-2016 Colorado Six-factor Multi-tiered System of Supports model; TDSL= Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; Layered Continuum of Supports; EBP= Evidence-Based Practice.

Specifically, results for this measurement model were χ^2 (1268, N =510) = 2743.69, $p < .01$; RMSEA = .05; CFI = .96; TLI = .96. The standardized path coefficient between TCAP and DS (β = -.06) was negative, small, and not statistically significant at the .01 level. This indicated district size did not explain a significant amount of the variance in 2014 third grade TCAP reading scores. With the addition of the DS exogenous variable, the standardized path between TCAP and MTSS, while small (β =

.18), remained positive and statistically significant. An examination of the range of λ values for the observed items and their individual factors revealed each item had a high factor loading value. Results are summarized in Table 35 and displayed in Appendix X.

Table 35

Six-Factor Higher-Order Structural Model with District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .92*	.54 - .85*
DBPS	7-9, 16-21, 23-25, 29, 32	.73 - .89*	.53 - .80*
USPM	40-50	.60 - .92*	.35 - .84*
LCS	27, 28, 31	.64 - .85*	.40 - .72*
FSCP	37-39	.83 - .91*	.69 - .82*
EBP	26, 30, 33-35	.83 - .95*	.69 - .90*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Five-Factor Higher-Order Structural Model with District Size

Next, the fit of the data to the original five-factor higher-order MTSS framework that used the expanded DBPS latent variable and included the highly correlated LCS latent variable in the hypothesized model with DS as an additional exogenous variable was conducted using SEM techniques (see Figure 18).

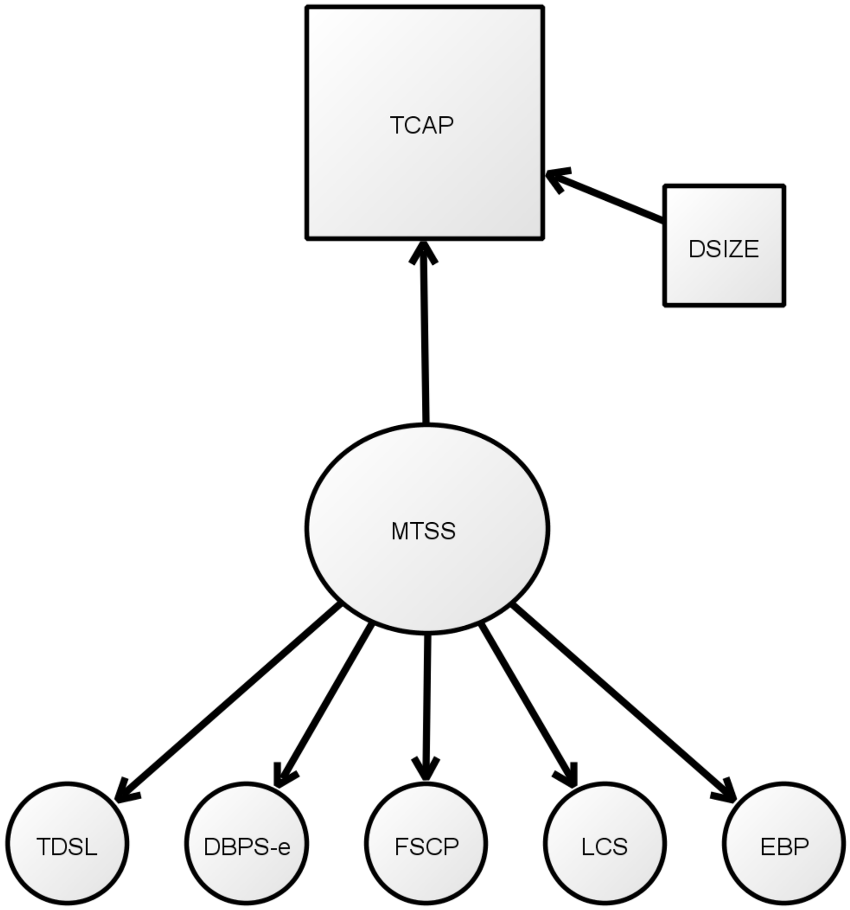


Figure 18. Five-factor higher-order structural model with district size.
Note. TCAP = 2014 third grade TCAP reading scores; DSize= Standardized K-12 enrollment for each school's district; MTSS = Current Colorado Multi-tiered System of Supports framework; TDSL= Team-Driven Shared Leadership; DBPS-e = Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; Layered Continuum of Supports; EBP= Evidence-Based Practice.

Results for this structural model indicated the six-factor model provided a better fit to the data: χ^2 (1269, N =510) = 3081.25, p < .01; RMSEA = .05; CFI = .95; TLI = .95. The path between TCAP and DS (β = -.06) remained negative, small, and statistically insignificant at the .01 level. This indicated district size did not explain a significant amount of the variance in 2014 third grade TCAP reading scores. However, with the addition of the DS exogenous variable, the standardized path between TCAP and MTSS, while small ($\beta = .18$), remained positive and statistically significant. An examination of the λ values for each observed item and the individual factors revealed each item had a high factor loading value. The ranges for each standardized indicator by latent factor, as well as ranges for SMC values, are included in Table 36 (see Appendix Y for the graphic representation).

Table 36

Five-Factor Higher-Order Structural Model with District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .92*	.54 - .86*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53 - .86*	.28 - .78*
LCS	27, 28, 31	.63 - .85*	.40 - .72*
FSCP	37-39	.83 - .91*	.69 - .82*
EBP	26, 30, 33-35	.83 - .95*	.69 - .91*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Revised Five-Factor Higher-Order Structural Model with District Size

Next, the fit of the data to the revised five-factor model that included both the DBPS and USPM latent variables and excluded the highly correlated LCS latent variable in the hypothesized model with DS as an additional exogenous variable was conducted using SEM techniques (see Figure 19).

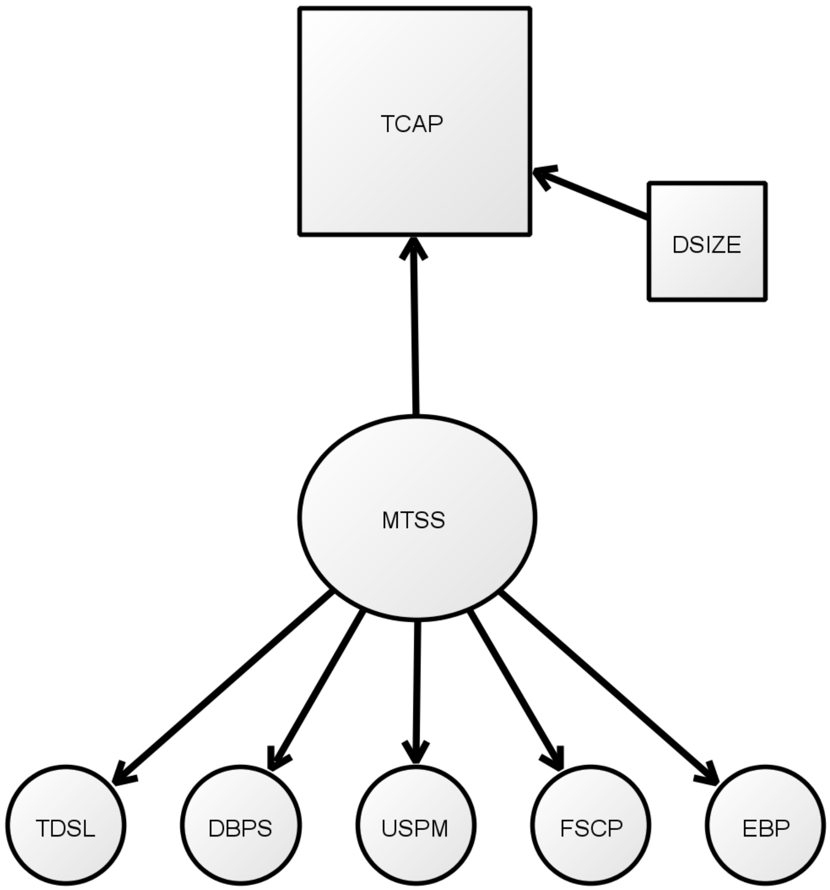


Figure 19. Revised five-factor higher-order structural model with district size.
Note. TCAP = 2014 third grade TCAP reading scores; DSize= Standardized K-12 enrollment for each school's district; MTSS = Pre-2016 Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-Driven Shared Leadership; DBPS= Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

This revised five-factor higher-order SEM, which did not include the highly correlated LCS latent endogenous factor, provided a better level of fit for the data than the previous five-factor model: χ^2 (1075, N = 511) = 2520.68, p < .01; RMSEA = .05; CFI = .96; TLI = .96. When compared with the previous model, the factor loadings differed from the original five-factor model in the data but were similar to the factor loadings of the six-factor model. The standardized path coefficient between DS and

TCAP continued to be statistically insignificant. However, the standardized path coefficient between MTSS and TCAP, while small. was also positive and statistically significant ($\beta = .18$). Results of the revised five-factor MTSS SEM are provided in Table 37and are represented graphically in Appendix Z.

Table 37

Revised Five-Factor Higher-Order Structural Model with District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74-.92*	.55 - .86*
DBPS	7-9, 16-21, 23-25, 29, 32	.73-.89*	.53 - .80*
USPM	40-50	.60-.92*	.36 - .84*
FSCP	37-39	.83-.91*	.70 - .82*
EBP	26, 30, 33-35	.83-.93*	.68 - .87*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Four–Factor Higher-Order Structural Model with District Size

The final higher-order structural model that examined the fit of the data to a model that contained the expanded DBPS factor but did not include the LCS factor and included DS as an exogenous variable provided an adequate fit for the data but did not exceed the level of fit for the revised five-factor or original six-factor higher-order SEMs: χ^2 (1123, N = 508) = 2828.97, p <.01; RMSEA = .06; CFI = .95; TLI = .95. Figure 20 provides a graphic representation of the four-factor model.

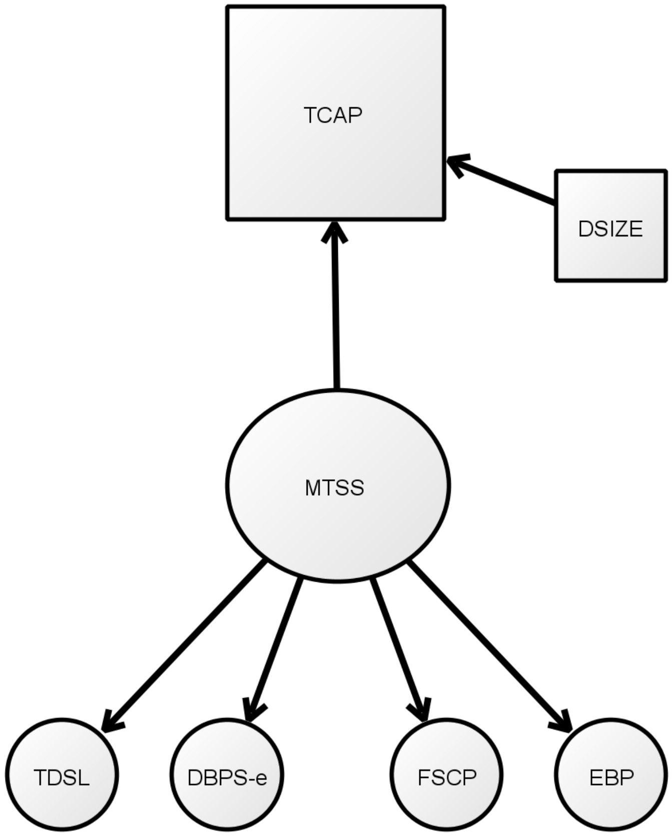


Figure 20. Four-factor higher-order structural model with district size.
Note. TCAP = 2014 third grade TCAP reading scores; DSize= Standardized K-12 enrollment for each school's district; MTSS = Current Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-Driven Shared Leadership; DBPS-e= Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

Similar to the previous structural models that included DS as an exogenous variable, the standardized path coefficient between MTSS and DS was not statistically significant. However, the small value of the standardized path coefficient between MTSS and TCAP was statistically significant (β = .18). Results of the analysis are summarized in Table 38 (see Appendix AA).

Table 38

Four-Factor Higher-Order Structural Model with District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .93*	.55 - .86*
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.54 - .89*	.52 - .79*
FSCP	37-39	.83 - .91*	.29 - .69*
EBP	26, 30, 33-35	.83 - .93*	.69 - .82*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Comparison of the Higher-Order Structural Models with District Size

The results from this set of higher-order structural models that included DS as an exogenous variable once again provided evidence the revised five-factor higher order MTSS model provided the best fit for the MTSS-IPS data. Specifically, RMSEA values for the revised five-factor model (.05), the CFI values (.96), and the normed TLI were identical to those generated for the six-factor model. However, the small value of the χ^2 statistic indicated the revised five-factor model provided a better fit for the data than any of the other proposed models. Table 39 provides a comparison of the results for the four models.

Table 39

Comparison of Higher-Order Structural Models with District Size

SEM	LV	*df*	χ^2	RMSEA	CFI	TLI
Six-factor	TDSL, DBPS, USPM, EBP, FSCP, LCS	1268	2743.69*	.05	.96	.96
Five-factor	TDSL, DBPS-e, EBP, FSCP, LCS	1269	3081.25*	.05	.95	.95
Five-factor (revised)	TDSL, DBPS, USPM, EBP, FSCP	1122	2520.68*	.05	.96	.96
Four-factor	TDSL, DBPS-e, EBP, FSCP	1123	2866.24*	.06	.95	.95

Note. Six-factor = Pre-2016 Colorado Six-factor MTSS Model; Five-factor = Post 2016 Colorado Five-factor MTSS model; Five-factor Revised = Six-factor without the LCS factor; Four-factor = Five-factor without the LCS factor; LV= Individual latent variables included in the model; *df*= Degrees of Freedom; χ^2= Chi-square test of model fit; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index; TLI= Tucker Lewis Index.
* $p < .01$

Final Models: Higher-Order Structural Models with District Size and Free and Reduced Lunch

Having tested the impact that adding district size as an exogenous variable on the fit of the data to a series of hypothesized higher-order six-factor MTSS models and determining in all cases that the revised five-factor model provided a better fit to the data, a final series of higher-order SEMs was conducted to investigate the impact that adding both DS and FRL as exogenous variables had on the fit of the data to the models. In the discussion that follows, the results of each model are presented.

Six-Factor Higher-Order Structural Model with District Size and Free and Reduced Lunch

To examine the fit of the data to the higher-order full structural model that included school-level 2014 third grade TCAP reading scores as the endogenous variable of interest and both DS and FRL as additional observed exogenous variables, the initial fit of the data to the original six-factor higher-order MTSS framework that included the non-positive definite LCS latent factor was conducted (see Figure 21).

Similar to the previous six-factor models, this model provided a good fit for the data: χ^2 (1318, *N* = 507) = 2643.20, *p* <.01; RMSEA = .05; CFI = .97; TLI = .97. The fit of the data to this higher-order six-factor model is summarized in Table 40. The standardized path coefficient between TCAP and FRL (β = -.69) was statistically significant and negative, which indicated the 2014 third grade TCAP reading scores tended to decrease as the percentage of students who qualified for free and reduced lunch increased. The standardized path coefficient between TCAP and DS, while smaller than the path coefficient between FRL and TCAP, was also negative and statistically significant (β = -.13). However, even with the addition of the statistically significant FRL and DS exogenous variables, the standardized path coefficient between TCAP and MTSS, while small, remained positive and statistically significant (β = .12). An examination of the λ values for each observed item and the individual factors revealed each item had a high loading factor value. The ranges for each standardized indicator by latent factor as well as ranges for SMC values are included in Table 40 (see Appendix AB for the graphic representation).

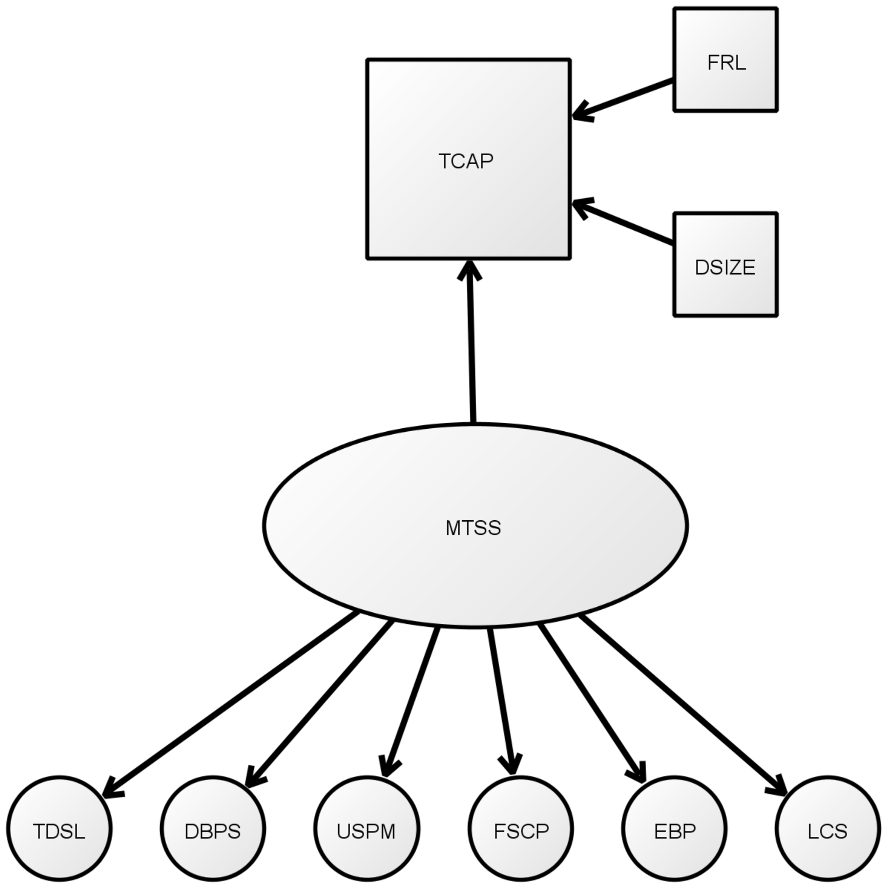


Figure 21. Six-factor higher-order structural model with free and reduced lunch and district size.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; DSize= Standardized K-12 enrollment for each school's district; MTSS = Pre-2016 Colorado Multi-tiered System of Supports Model; TDSL= Team-Driven Shared Leadership; DBPS = Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice; LCS= Layered Continuum of Supports.

Table 40

Six-Factor Higher-Order Structural Model with Free and Reduced Lunch and District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74 - .93*	.54 - .87*
DBPS	7-9, 16-21, 23-25, 29, 32	.73 - .90*	.53 - .80*
USPM	40-50	.59 - .92*	.35 - .84*
LCS	27, 28, 31	.64 - .85*	.41 - .73*
FSCP	37-39	.83 - .91*	.70 - .83*
EBP	26, 30, 33-35	.83 - .96*	.68 - .92*

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Five-Factor Higher-Order Structural Model with District Size and Free and Reduced Lunch

Next, the initial fit of the data to the original five-factor higher-order MTSS framework that included both FRL and DS as exogenous variables was conducted using SEM techniques (see Figure 22).

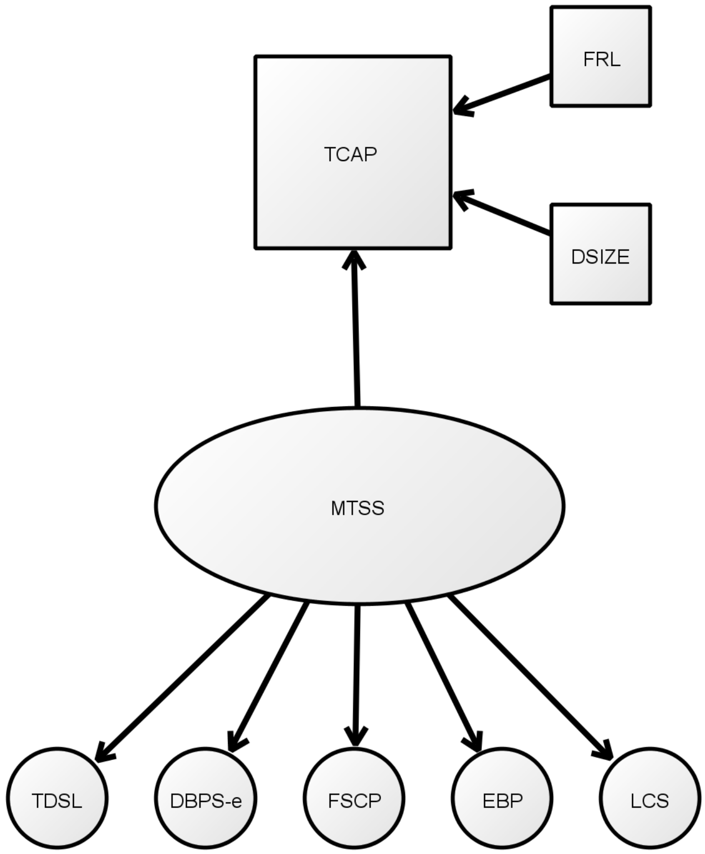


Figure 22. Five-factor higher-order structural model with free and reduced lunch and district size.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; DSize= Standardized K-12 enrollment for each school's district; MTSS = Current Colorado Multi-tiered System of Supports framework; TDSL= Team-Driven Shared Leadership; DBPS-e = Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice; LCS= Layered Continuum of Supports.

As in the previous original five-factor models, this model included the expanded DBPS latent factor as well as the troublesome, highly correlated, and non-positive definite LCS latent endogenous factor. The global fit indices generated from this analysis demonstrated that again the data provided good fit to the model. Specifically, results for this five-factor measurement model were χ^2 (1319, N =507) = 2965.76, p < .01; RMSEA

= .05; CFI = .96; TLI = .96. Similar to the previous six-factor higher-order SEM that included FRL and DS as exogenous variables, the path coefficient between FRL and TCAP was negative and statistically significant (β = -.69). The standardized path coefficient between TCAP and DS, while smaller than the path coefficient between FRL and TCAP, was also negative and statistically significant (β = -.13). The small standardized path coefficient between TCAP and MTSS also continued to be statistically significant (β = .13). An examination of the λ values for each observed item and the individual factors revealed each item had a high loading factor value. The ranges for each standardized indicator by latent factor as well as ranges for SMC values are included in Table 41 (see Appendix AC for the graphic representation).

Table 41

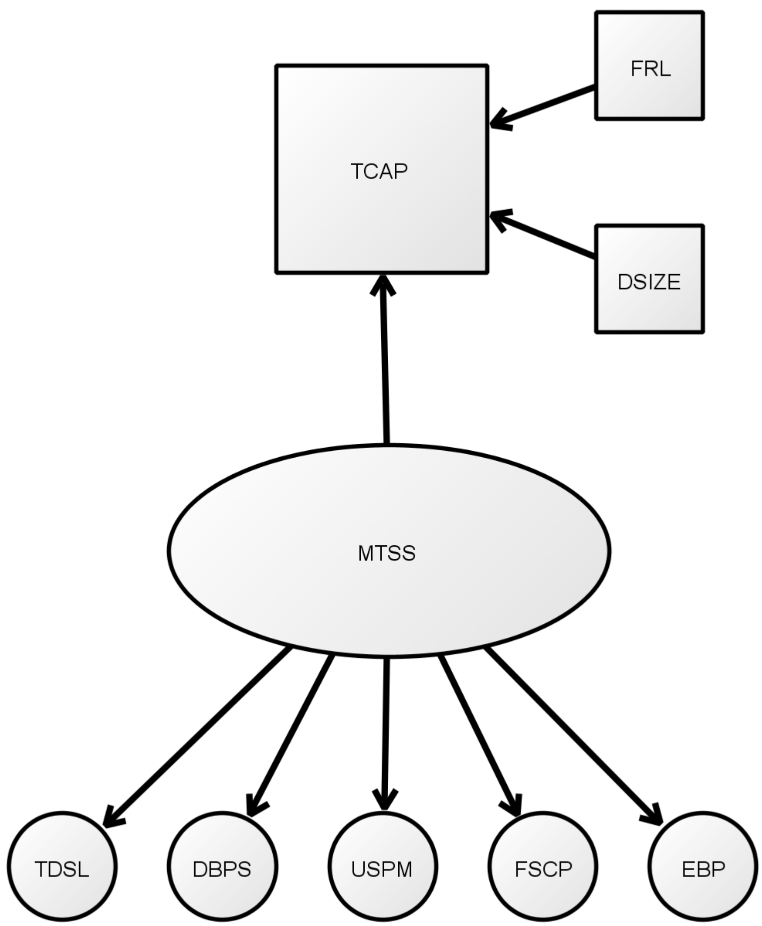
Five-Factor Higher-Order Structural Model with Free and Reduced Lunch and District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74-.93*	.54-.87
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.53-.89*	.52-.78
LCS	27, 28, 31	.64-.85*	.41-.73
FSCP	37-39	.83-.91*	.70-.82
EBP	26, 30, 33-35	.83-.96*	.68-.92

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Revised Five-Factor Higher-Order Structural Model with District Size and Free and Reduced Lunch

Next, the fit of the data to the revised five-factor model that included the USPM and DBPS factors but eliminated the LCS factor was assessed (see Figure 23).



.

Figure 23. Revised five-factor higher-order structural model with free and reduced lunch and district size.
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; DSize= Standardized K-12 enrollment for each school's district; MTSS = Pre-2016 Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-Driven Shared Leadership; DBPS= Data-Based Problem Solving; USPM = Universal Screening and Progress Monitoring; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

Similar to the previous revised five-factor higher-order models both with and without additional exogenous variables, this revised five-factor model also provided an improved fit for the data over the previous similar models. Specifically, the statistics and fit indices were χ^2 (1169, N =507) = 2408.14, p < .01; RMSEA = .05; CFI = .97; TLI = .97. As in the previous structural models that included FRL and DS as exogenous variables, the standardized path coefficients between TCAP and FRL was large, negative, and statistically significant (β = -.69) while the path between TCAP and DS remained small but statistically significant (β = .13). Finally, the standardized path coefficients between MTSS and TCAP remained small but statistically significant. Results are summarized in Table 42 and represented graphically in Appendix AD.

Table 42

Revised Five-Factor Higher-Order Structural Model with Free and Reduced Lunch and District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.76-.93*	.55-.87
DBPS	7-9, 16-21, 23-25, 29, 32	.74-.90*	.53-.80
USPM	40-50	.72-.92*	.35-.85
FSCP	37-39	.83-.91*	.70-.83
EBP	26, 30, 33-35	.84-.94*	.68-.88

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* p < .01.

Four-Factor Higher-Order Structural Model with District Size and Free and Reduced Lunch

The final higher-order structural model that examined the fit of data to a model that contained the expanded DBPS factor but did not include the LCS factor provided an adequate fit for the data; however, it did not exceed the level of fit for the revised five-factor or original six-factor higher-order SEM. Figure 24 provides the reader with a graphic representation of the four-factor model.

As before, the fit of the data to a final four-factor model that included the expanded DBPS factor, excluded the LCS factor, and included the exogenous FRL and DS exogenous indicators was examined. Results indicated the model did not provide a better fit for the data than the previous revised five-factor model: χ^2 (1170, N = 507) = 2737.00, p <.01; RMSEA = .06; CFI = .96; TLI = .96. Similar to the previous structural models that included FRL and DS as exogenous variables, the standardized path coefficients between TCAP and FRL was large, negative, and statistically significant (β = -.69) while the standardized path coefficient between DS and TCAP was small but remained statistically significant (β = .12). As with previous similar models, the standardized path coefficient between TCAP and MTSS was small, positive, and statistically significant (β = .13). Results of the analysis are summarized in Table 43 and provided graphically in Appendix AE.

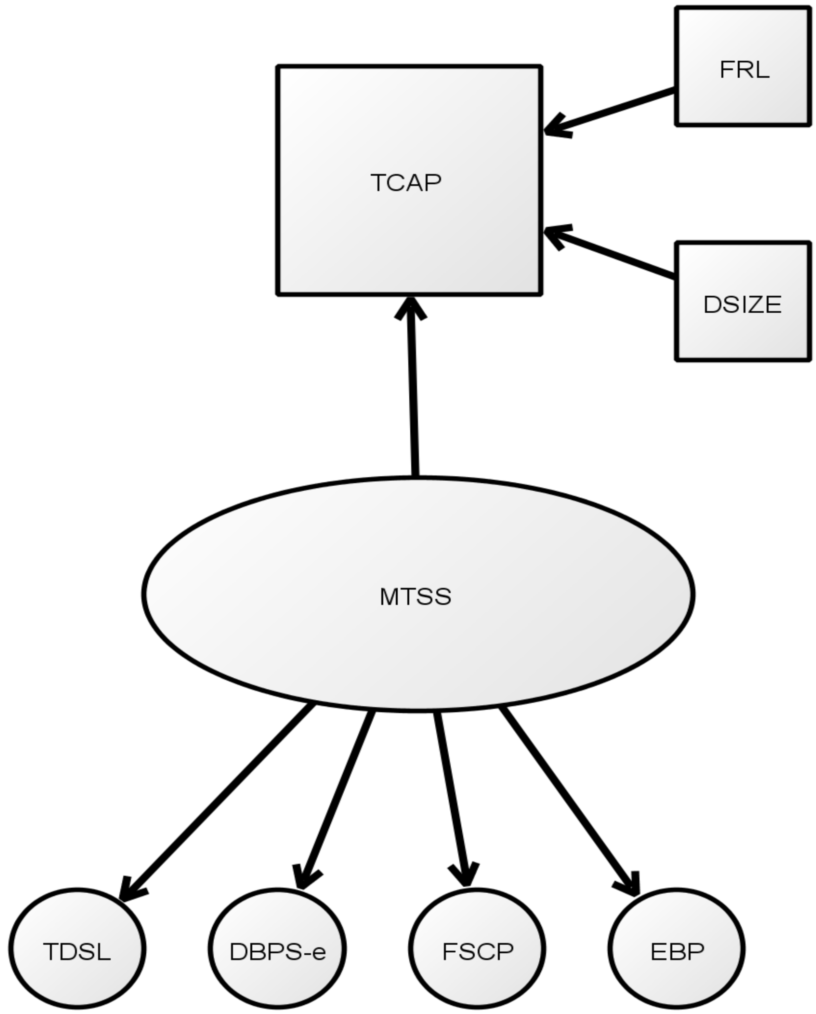


Figure 24. Four-factor higher-order structural model with free and reduced lunch and district size**.**
Note. TCAP = 2014 third grade TCAP reading scores; FRL = % of students in each school who qualified for the free and reduced lunch (FRL) program; DSize= Standardized K-12 enrollment for each school's district; MTSS = Current Colorado Multi-tiered System of Supports Framework without Layered Continuum of Supports; TDSL= Team-Driven Shared Leadership; DBPS-e= Expanded Data-Based Problem Solving; FSCP= Family, School, and Community Partnerships; EBP= Evidence-Based Practice.

Table 43

Four-Factor Higher-Order Structural Model with Free and Reduced Lunch and District Size

LV	IPS Items	λ (Range)	SMC (Range)
TDSL	1-6, 10-15, 22, 36	.74-93*	.54-.87
DBPS-e	7-9, 16-21, 23-25, 29, 32, 40-50	.67-.89*	.28-.79
FSCP	37-39	.83-.91*	.70-.83
EBP	26, 30, 33-35	.83-.94*	.68-.88

Note. LV= Latent Variable; IPS Items = Items included in the latent variable; λ (Range) = Range of Completely Standardized Factor Loading for the LV items; SMC (Range) = Range of Item-specific Squared Multiple Correlations.
* $p < .01$.

Comparison of the Higher-Order Structural Models with District Size and Free and Reduced Lunch

The results from this set of higher-order structural models that included both FRL and DS as exogenous variables provided evidence the revised five-factor higher order hypothesized MTSS model provided the best fit for the MTSS-IPS data. Specifically, RMSEA values for the revised five-factor model (.05), the CFI values (.97), and the normed TLI (.97) were identical to those generated for the six-factor model. However, the smaller value of the χ^2 statistic indicated the revised five-factor model provided a better fit for the data than any of the other proposed models. Table 44 provides a comparison of the results for the four models.

Table 44

Comparison of Higher-Order Structural Models with Free and Reduced Lunch and District Size

SEM	LV	*df*	χ^2	RMSEA	CFI	TLI
Six-factor	TDSL, DBPS, USPM, EBP, FSCP, LCS	1318	2643.20*	.05	.97	.97
Five-factor	TDSL, DBPS-e, EBP, FSCP, LCS	1319	2965.76*	.05	.96	.96
Five-factor (revised)	TDSL, DBPS, USPM, EBP, FSCP	1169	2408.14*	.05	.97	.97
Four-factor	TDSL, DBPS-e, EBP, FSCP	1170	2737.00*	.05	.96	.96

Note. Six-factor = Pre-2016 Colorado Six-factor MTSS Model; Five-factor = Post 2016 Colorado Five-factor MTSS model; Five-factor Revised = Six-factor without the LCS factor; Four-factor = Five-factor without the LCS factor; LV= Individual latent variables included in the model; *df*= Degrees of Freedom; χ^2= Chi-square test of model fit; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index; TLI= Tucker Lewis Index.
* $p < .01$.

Direct Effects of the Latent Factors with Third Grade Reading

An examination of the results revealed various direct effects between the latent factors of the six-component MTSS framework and the 2014 third grade TCAP reading scores were statistically significant (see Appendix AF). The standardized path coefficient between the DBPS latent variable and TCAP was small but positive and statistically significant (β = .11, SE = .04, *p*-value < .01). This indicated that as perceptions of data-based problem solving processes increased, 2014 third grade reading scores also increased. Similarly, the standardized path coefficient between the USPM latent variable and TCAP scores was also small, positive, and statistically significant (β- .12, SE = .05, *p*-value < .01). This indicated that higher perceptions of universal screening and progress monitoring were predictive of higher TCAP scores. Higher perceptions associated with evidence-based practices were also predictive of higher TCAP scores (β = .15, SE = .05, *p*-value < .01). When participants reported higher perceptions of coaching and professional development opportunities combined with stronger perceived classroom and behavior management capabilities, TCAP scores tended to improve. Comparable results were obtained for perceptions of family, community, and school-level partnerships (β = .194, SE = .05, *p*-value < .01). Finally, the results revealed the participants' perceptions associated with TDSL, while positive, were not statistically significant (β = .07, SE .04, *p*-value > .01). This indicated perceptions of leadership not part of a systemic framework were not predictive of higher TCAP scores.

As noted in earlier discussion, an examination of the direct effects of the MTSS systemic framework revealed perceptions of MTSS implementation were predictive of increased 2014 third grade TCAP reading scores when both FRL and district size were included as moderating variables (β = .12, SE= .04, *p*-value < .01). However, FRL remained negatively related to TCAP scores (β = -.69, SE =.02, *p*-value < .01). In other words, as 2014 third grade reading TCAP scores increased, the percentage of students within each school who qualified for FRL decreased. Additionally, district size was also statistically significant and negatively related to 2014 third grade reading TCAP scores (β = -.13, SE = .04, *p*-value < .01). Larger district size was predictive of lower TCAP scores.

In conclusion, the preliminary data analysis did not reveal any noteworthy measurement flaws or threats to the generalizability of the results of this study. A series of confirmatory factor analyses were conducted on the individual latent factors and on a series of models. While the individual factor analyses indicated the separate factors provided a good level of fit for the observed items, the higher-order confirmatory factor analyses and SEMs revealed the LCS latent factor was highly correlated with a variety of the other latent factors. Subsequent analyses revealed when the LCS latent factor was removed from the models, the fit of the data typically improved. An examination of the standardized path coefficients between the endogenous variables with both MTSS and TCAP revealed the MTSS remained positive, statistically significant, and predictive of higher 2014 third grade TCAP reading scores. The next chapter develops some of the ideas presented in this chapter and discusses the implications of the study for the field.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The previous four chapters provided an introduction and rationale for the study, summarized the body of literature that grounded this investigation, detailed the methodology and statistical model that were used, and provided the results of those analyses. Specifically, the review of MTSS-related experimental, quasi-experimental, and qualitative research studies that investigated elements of the MTSS framework was used to develop a series of hypothetical higher-order structural models that examined the relationship between the components of a MTSS model and the correlation they had with third grade reading achievement. To investigate this relationship, perception of MTSS implementation data served as the independent variable and publicly available 2014 third grade TCAP reading scores were used as the dependent variable of primary interest. After providing a brief summary of the participants' demographics, this final chapter presents a summary and discussion of the study's research findings and highlights some of the more significant results. The implications of the findings are provided and are followed with a discussion of the study's limitations. The chapter concludes with recommendations for future studies.

Demographic Characteristics

This study appeared to have captured the perceptions of MTSS implementation from a representative sample of Colorado educators. The MTSS-IPS was distributed to approximately 1,500 individuals employed in the state of Colorado with a range of educator roles. Individuals with target roles (e.g., third grade teachers, special education resource teachers, elementary school principals) or their proxies were randomly identified at the school level; the MTSS-IPS had a 25.06% response rate. Of the responders, 86.47% were female and 13.52% were male. These results were somewhat equivalent to demographic information from the 2014-2015 academic year provided by the CDE (2014). In 2014-2015, approximately 75% of all Colorado teachers were women (n = 39,859). Participants worked within one of 306 individual schools that were distributed across a vast majority of Colorado school districts (72.68%). A majority of the responders were third grade teachers (38.96%) while the remaining responders either served students as administrators (12.75%), special education resource teachers (27.58%), or had a similar proxy role (20.68%).

Data about the participants' years of experience were gathered in this study and a majority (64.43%) of the participants shared they had been a licensed educator in excess of 10 years. Josephson (2015) shared that a recent Gallup poll found the average retirement age of teachers hovered around 59, which meant teachers were electing to stay in the profession longer. Because a majority of the participants in this study had also been working with students for more than 10 years, it is likely their perceptions were comparable to other educators in Colorado in general and the nation at-large.

Discussion of Findings

In the following discussion, a general explanation of the results and a detailed discussion of the notable highlights are provided. To review, structural equation modeling (SEM) was used to provide an understanding of the MTSS models adopted in the state of Colorado and how educators' perceptions of implementation correlated with 2014 third grade TCAP reading scores. A series of increasingly complex confirmatory factor analyses and structural equation models were tested; therefore, a discussion that detailed how MTSS implementation perceptions related to 2014 third grade TCAP reading data was required.

The major focus of this study was to test the fit of a series of higher-order SEM models that examined how perceptions of MTSS implementation correlated to 2014 third grade TCAP reading outcomes. McIntosh and Goodman (2016) shared the MTSS framework evolved out of the RTI and PBIS initiatives and strove to provide teachers with technical assistance and professional support as they worked to meet the academic and behavioral needs of students. However, there is a lack of national consensus on the critical components that should be included in an MTSS framework and how those components should be defined (Samuels, 2016). Additionally, a variety of experts have shared that the initiative is complex and could be difficult to implement (Balu et al., 2015; Hudson, 2013).

While the novelty of the MTSS initiative might mean more time is needed for the field to develop a clear understanding of the important components of an effective MTSS model, time is a luxury striving readers simply cannot afford. Previous research consistently demonstrated that when students fail to read proficiently early during their education, that failure resulted in significant, long-term, and negative consequences that impacted those students as individuals, their families, and the nation at large (Fiester, 2010; Planty et al., 2008). Therefore, this study sought to investigate how perceptions of MTSS implementation in Colorado correlated with 2014 third grade reading outcomes and examined whether increased implementation perceptions correlated with improved third grade reading outcomes. In the discussion that follows, key findings are discussed within the context of the research questions.

Research Question 1(a)

The primary research question posed in this study asked if the hypothesized higher-order MTSS theoretical factor structure of each measurement model fit the data. To answer this question, a series of CFAs were conducted to assess the fit of the data to the individual latent constructs of the hypothesized measurement models.

Confirmatory factor analysis: Team-driven shared leadership. The TDSL factor survey items primarily asked participants to evaluate how the leaders within their school (a) were committed to increasing student learning outcomes, (b) actively engaged with parents and teachers, (c) encouraged collaboration between staff and with families, and (d) provided training and resources to increase teachers' pedagogical competencies. Results revealed the TDSL latent factor provided a good fit for the data. Individual item factor loadings, SMC values, and the CFI and TLI fit indices all exceeded the minimum values. For example, the factor loadings for the observed indicators of the TSDL factor ranged between .73 and .91(see Table 12 and Appendix E). According to Yong and Pearce (2013), because factor loadings are an indicator of how much an observed indicator adds to the factor of interest, loadings smaller than .30 might indicate a weak relationship. The results indicated the TDSL latent factor provided a good fit for the data.

Confirmatory factor analysis: Universal-screening and progress monitoring. In an MTSS framework, data were collected, studied, summarized, and employed (a) to assess the fidelity of implementation, (b) for diagnostic and screening purposes, (c) to monitor student progress, and (d) to inform general outcomes at the school-level (McIntosh, Reinke, & Herman, 2009; Torgesen, 2009). Each source of data could be used to improve the learning outcomes of all students. Therefore, universal screening and progress monitoring survey items asked participants to share details about the structure and systems used within their schools to gather, store, access, and interpret a range of academic and behavioral data for all students based on the work of previous researchers (e.g., Chard et al., 2008; Fuchs et al., 2008; McIntosh & Goodman, 2016; Wanzek & Vaughn, 2007). The CFA results revealed the measurement model of the 11-item USPM latent factor also provided an adequate fit for the data. While the RMSEA, CLI, and TLI indices fell slightly outside of the ideal range, the individual factor loadings ranged between .61and .90; all were significant at the $p < .01$ level (see Table 14 and Appendix G). Therefore, it could be confidently concluded the USPM latent factor meaningfully contributed to the MTSS measurement model.

Confirmatory factor analysis: Data-based problem solving. Multi-tiered systems like MTSS are primarily data-driven (McIntosh & Goodman, 2016). Student screening data and progress monitoring data are used to inform both instruction and the scale-up efforts of the system as a whole. While using and interpreting data could be intimidating for many educators, having positive experiences with data could help alleviate those fears (McIntosh & Goodman, 2016). Facilitating data interpretation takes a problem-solving approach where teams focus on interpreting the information so it can be used in a real and meaningful way (McIntosh & Goodman, 2016). The 14 MTSS-IPS items theorized to load on the DBPS factor addressed concepts associated with analyzing, interpreting, and using data to drive their instruction. Results demonstrated the DBPS latent endogenous factor provided a good level of fit for the data (see Table 13 and Appendix F). Individual factor loadings of the observed items for the DBPS latent factor ranged between .70 and .91. While the RMSEA value of .14 was above the ideal cut-point of .06, the values for the CLI and TLI indices were .97 and .96, respectively, which met the minimum value for categorical data (Schreiber et al., 2006).

Confirmatory factor analysis: Expanded data-based problem solving. There was enough evidence to also conclude the expanded 25-item DBPS factor also provided adequate fit for the data. To review, the expanded 25-item DBPS factor combined the 11 items from the USPM factor with the 14 items of the DBPS latent factor. The theoretical justification for this combination was based on recent work of leaders in the CDE (2016) who wanted to create a more efficient MTSS model that could be brought to scale throughout the state of Colorado. Confirmatory factor analysis results of the expanded DBPS latent factor were similar to those generated by the individual CFAs of the two individual elements (i.e., DBPS and USPM). Recalling information provided in Chapter III, Cronbach α reliability estimates of the expanded DBPS were .96; therefore, it provided reasonable evidence the 25 items generated consistent data (see Table 8). While the values of the fit indices fell slightly outside of the recommended range (RMSEA = .11; CFI = .93; TLI = .93), each of the indicator items ranged between .55 and .90; the SMC values ranged between .31 and .80, which indicated an adequate fit of the data to the measurement model of the expanded DBPS latent factor (see Table 20 and Appendix L).

Confirmatory factor analysis: Evidence-based practice. Because MTSS scale-up efforts are used to improve the learning outcomes of students, using instructional strategies that research demonstrates increases student learning outcomes is critical (Al Otaiba et al., 2016; Menzies et al., 2008; Rodriguez & Denti, 2011). Survey items theorized to load on the EBP factor asked participants to evaluate the systems used within their schools to provide technical assistance and support for teachers as they learned how to incorporate evidence-based instructional strategies, classroom management routines, and behavioral supports in their work with students. The results of the CFA for the EBP latent factor indicated the measurement model provided a good fit for the data (see Table 17 and Appendix J). Individual factor loadings for the observed items ranged between an acceptable .59 and .96 (Tabachnick & Fidell, 2007). While the value for the RMSEA fit index was higher than the ideal (.23), the CFI and TLI values were .98 and .97, respectively, which fell above the recommended .96 (Schreiber et al., 2006). Therefore, there was evidence the measurement model of the EBP factor contributed in a meaningful way to the overall model.

Confirmatory factor analysis: Layered continuum of supports. When viewed comprehensively, the survey items forced to load on the LCS latent factor addressed concepts associated with differentiated instructional strategies that moved from less intensive to more intensive as student needs increased. Reasonable evidence also indicated the LCS latent factor provided a good fit for the data (see Table 16 and Appendix I). However, results of the higher-order CFAs and SEMs that included the LCS latent factor revealed the factor was highly collinear and not positive definite (see

Table 18 and Appendix K). Ultimately, this highly collinear factor was removed from the final models in an effort to improve the fit of the data to the model.

Confirmatory factor analysis: Family, school, and community partnerships. Family members contributed and participated in the MTSS leadership team. Incorporating family perspectives on the team could serve to enrich family and school partnerships (Garbacz et al., 2016; Senechal & Young, 2008). Survey items related to the FSCP factor asked the participants to assess how they included and engaged with families in their child's learning. Reasonable evidence indicated the FSCP latent factor provided a good level of fit for the three observed indicator variables. Examinations of the individual factor loadings revealed the λ values were all at or above an acceptable .65 level (see Table 15 and Appendix H). As a result, it could also be confidently stated the FSCP latent factor contributed in a meaningful way to the MTSS measurement model.

Higher-order confirmatory factor analyses. Having determined the individual factors provided a reasonable fit for the MTSS-IPS data, higher-order CFAs were conducted to determine if one of the MTSS models would provide a better fit for the data than the others. Results indicated the revised five-factor MTSS model provided a more accurate representation of perceptions of MTSS implementation than any of the other proposed models. Specifically, the revised five-factor model that included (a) TDSL, (b) DBPS, (c) USPM, (d) EBP, and (e) FSCP latent factors provided the best fit for MTSS-IPS data. Further, the completely standardized path coefficients among each of the five endogenous latent factors were all large, positive, and statistically significant. This indicated that comprehensively, perceptions associated with leadership, data, evidence-based practices, and family partnerships were predictive of increased perceptions of MTSS implementation.

Research Question 1(b)

A current gap exists in the body of research that investigates how educators' perceptions of implementing a comprehensive MTSS framework correlate with student reading outcomes. Therefore, MTSS-IPS data were fit to a series of higher-order structural equation models to clarify the relationship between perceptions of MTSS implementation and 2014 third grade TCAP reading outcomes. All the tested models used a variety of endogenous models that differed by the number of endogenous latent factors they contained. A discussion of the results and their implications follows.

Because the main purpose of this study was to examine how perceptions of MTSS implementation related to 2014 reading outcomes, a second set of higher-order SEMs was conducted (see Figures 5-8 and Appendices P-S). Similar to the previous example, the fit of the data to the four proposed models was compared. Results revealed when 2014 third grade TCAP scores were added as a higher-order endogenous variable to the models, the revised five-factor MTSS model provided a better fit for the data than the six-factor, original-five factor, or four-factor models (Table 29 and Appendix Q). Notably, the standardized path coefficient that examined how perceptions of MTSS implementation related to 2014 third grade TCAP reading scores, while small, was positive and statistically significant ($\beta = .18$, SE = .06, $p < .01$). These results suggested that when model efficiency is the ultimate goal, including elements associated with (a) leadership, (b) gathering data, (c) using data in a problem solving process, (d) evidence-based instructional practices, and (f) family, school, and community partnerships should be included.

Research Question 1(c)

Information shared in previous chapters demonstrated the long-term negative effects reading below proficient levels have on students. Briefly, previous researchers have convincingly demonstrated that when students fail to learn early in their educational career, they learn less (e.g., Cunningham & Stanovich, 1997), have lower levels of self-esteem (e.g., Rose, 2006), are more likely to drop out of school (e.g., Compton et al., 2012), and on average earn $25,000 less per year than their peers with a college degree (Planty et al., 2008). Further, Byrd (2005) shared that adults who do not have a high school education are also more likely to exhibit health-impairing behaviors, to suffer from poor health, and to die at younger ages than their peers who successfully graduated from high school. Put simply, children suffer for the duration of their lives when they fail to learn to read.

To begin to explain how perceptions of MTSS implementation related to third grade reading outcomes and account for the variance in the model that could be attributed to schools who serve students from families who are economically disadvantaged, the percentage of students within each school who qualified for federally funded free or reduced school lunch (FRL) was included as an exogenous variable in a second set of higher-order SEMs. To explain, students whose household incomes fell below the poverty level were provided with free and reduced-price meals by the federal government (U.S. Department of Agriculture, 2016; see Figures 13-16 and Appendices T-W). When the level of fit for higher-order MTSS models that included the exogenous FRL variable was compared, the results again revealed the revised five-factor model provided a slightly better fit. Specifically, the revised higher-order five-factor model generated an RMSEA value of .05 and CFI and TLI values of .96. Factor loadings for the individual latent constructs of the model (e.g., TDSL, DBPS, USPM, EBP, FSCP) had a lower limit of .59 and an upper limit of .94; all were significant at the .01 level (see Table 32 and Appendix V). The standardized path coefficient between FRL and TCAP was large, negative, and statistically significant ($\beta = -.67$, SE = .04, $p < .01$). However, the standardized path coefficient that examined how perceptions of MTSS implementation related to 2014 third grade TCAP reading scores when FRL was included as an exogenous variable, while smaller than previous results, it was positive and remained statistically significant ($\beta = .12$, SE = .04, $p < .01$). These results indicated increased perceptions of MTSS implementation were predictive of higher third grade TCAP reading scores even when students were impacted by poverty.

Research Question 1(d)

According to the U.S. Department of Education (USDOE; 2016), a majority of the funding for public education is generated by individual states and local communities who use financial resources generated by taxes. For example, of the estimated $1.15 trillion spent throughout America during the 2012-2013 academic year, approximately 92% of the financial resources needed to fund education came from either state or local tax revenue streams (USDOE, 2016). Given that larger districts have a larger tax base, it is reasonable to expect those districts would have more financial resources to meet student needs and smaller districts would have access to less financial resources. Therefore, to investigate the effect district enrollment numbers had on 2014 third grade

TCAP reading scores, a standardized measure of district size was included as an exogenous variable on a second set of higher-order structural MTSS models (see Figures 17-20 and Appendices X-AA). Once again, the fit of the data to the revised five-factor higher-order structural model provided the best fit (see Table 37 and Appendix Z). Specifically, the revised five-factor model that included TDSL, DBPS, USPM, FSCP, and EBP produced an RMSEA value of .05 and CFI and TLI values of .96. When the standardized path coefficients between TCAP and both district size and MTSS were examined, the results revealed district size had a small, negative, and statistically insignificant impact on TCAP scores (β = -.06, SE = .04, p = .19). However, perceptions of MTSS while small, were positive and statistically significant even when the amount of tax funding each district received was included (β = .18, SE = .06, p < .01).

The final model tested for this study examined the fit of the data generated by the Colorado sample of participants to series of structural model that included both DS and FRL as exogenous variables (see Figures 21-24). Consistent with previous results, the revised five-factor model generated the closest fit over all previous models reported in this study (see Table 27 and Appendix AD). Specifically, the χ^2 value was 2408.14 (1318, N = 507) p < .01. The RMSEA pre-rounding value was .046, which was the smallest value for this statistic generated by any of the previous models. Similarly, the values of the CFI and TLI fit statistics value were .97--both .01 points higher than the minimum value for acceptable fit. An examination of the path coefficients that led from MTSS, which was the exogenous variable of primary interest, demonstrated that perceptions of MTSS implementation were positively related to 2014 third grade TCAP reading scores with a standardized coefficient (β) value of .12 and a standardized error

(SE) of .04 (see Appendix AD). An examination of the other path coefficients of the model demonstrated as expected that the percentage of students who qualified for free and reduced lunch was negatively related to 2014 third grade TCAP scores (β = -.69, SE .02). Surprisingly, in this model, the exogenous variable district size also became statistically significant and negatively related to third grade TCAP reading scores (β = -.13, SE = .04), although to a smaller degree than the percentage of student who qualified for the federal free and reduced lunch program. Finally, both district size and the percentage of students who qualified for the federal free and reduced lunch program appeared to vary jointly.

The path coefficients for the individual endogenous latent factors of the revised five-factor higher order model indicated perceptions of the latent factors were positively related to MTSS implementation perceptions (see Appendix AD). Specifically, perceptions of the use of evidence-based practices were large, statistically significant, and positively related to perception of MTSS implementation (β = .83, SE = .02) as were perceptions of family, community, and school partnerships (β = .80, SE = .03). Similarly, perceptions associated with gathering and using student-level data during the instructional planning process were also large, statistically significant, and positively related to MTSS implementation perceptions. Specifically, path coefficients between perceptions of universal screening and progress monitoring and MTSS implementation were positive (β = .84, SE = .02); there was a large positive relationship between data-based problem solving processes and MTSS implementation perceptions (β = .95, SE = .01). In other words, perceptions associated with gathering and using data were positively related to MTSS implementation perceptions. Similarly, perceptions of team-driven shared leadership were large and also strongly predictive of MTSS implementation perceptions (β = .94, SE = .01).

Research Question 1(e)

To investigate if any of the latent factors accounted for more of the variance in student reading outcomes than others when MTSS implementation was removed from the model, a final set of analyses was conducted. Direct effects evaluated by examining the paths from each of the six endogenous latent variables of the final model directly with the 2014 third grade reading variable were examined (see Appendix AF). Results revealed perceptions of family, school, and community partnerships were small but statistically significant and positively related to 2014 third grade reading outcomes (β = .19, SE = .05, $p < .01$). As in the previous example, when participants reported higher levels of engagement and interaction with families when their students' needs increased, student reading scores within the schools included in the sample also tended to increase to a statistically significant degree. Similarly, the path coefficient between TCAP and perceptions of evidence-based practices was small but statistically significant and predictive of higher third grade 2014 TCAP reading scores (β= .15, SE = .05, $p < .01$). Results also revealed the effect of perceptions of (a) universal screening and progress monitoring (β = .12, SE = .05, $p < .01$) and (b) data-based problem solving (β = .11, SE = .04, $p < .01$) were both small but also statistically significant and positively related to TCAP scores. As perceptions associated with data gathering and usage increased, TCAP scores also increased. When viewed comprehensively, these results suggested as practices associated with MTSS increase, student reading outcomes also tend to increase. However, the analysis of the results also indicated the direct effect of perceptions of leadership on TCAP scores was both small and statistically insignificant ($\beta = .07$, SE = .04, *p*-value = .07). This finding was contrary to previous results that indicated perceptions of team-driven shared leadership were large, a statistically significant predictor of MTSS perceptions, and contributed to increased reading outcomes for students. In other words, when leadership was not housed within an overarching system like the MTSS, increased perceptions of leadership did not have a large impact on student reading outcomes.

Additionally, because one of the secondary purposes of this study sought to investigate how the fit of the data generated by the MTSS-IPS compared a series of MTSS higher-order models, the results of an additional set of direct effects were examined to study the direct effects of the expanded 25-item DBPS factor on TCAP scores. To briefly review, the direct effect of the 11-item USPM factor and TCAP scores was small, positive, and statistically significant ($\beta = .12$, SE = .05, $p < .01$) as was the direct effect on the 14-item DBPS latent factor and TCAP scores ($\beta = .11$, SE = .04, $p < .01$). When the two factors were combined in the four-factor model, the direct effects between TCAP and the expanded DBPS factor were similar to the effects of the USPM factor alone ($\beta = .12$, SE= .04, $p < .01$). From these results, it could be inferred that when school-based professionals had higher perceptions of gathering, analyzing, and using student-level data to inform their instructional decision-making process and made decisions about students who might benefit from increased levels of support, student reading achievement increased. For many educators serving students in schools, the notion of using data could provoke anxiety and stress and could be a confusing and convoluted process (McIntosh & Goodman, 2016). However, results of this study indicated higher perceptions of data gathering and usage were correlated with higher student reading outcomes and, therefore, were important components of MTSS models.

Implications

When viewed comprehensively, the results indicated that when an MTSS framework included components associated with (a) leadership, (b) evidence-based instructional practices, (c) universal screening and progress monitoring, (d) data-based problem solving, and (e) partnerships between families and schools, student reading outcomes tended to improve. Since the MTSS framework is a school-based initiative created to provide teachers with support as they work to meet the needs of all their students, it stands to reason increased awareness of those supports would correlate with increased learning of students. In the discussion that follows, individual findings of this study are compared with those from previous research.

The Importance of Leadership

The results found that team-driven shared leadership became a critical component of MTSS implementation efforts. The structural and measurement components of the models included in this study supported the idea that teaming and leadership structures were important vehicles that facilitated student learning. These findings were similar to a variety of previous studies that explicitly examined the impact collaboration, professional development, and technical assistance had on student achievement and found these structures could be used to help advance student learning outcomes (Gil &Woodruff, 2011; Regan et al., 2015; Shepherd & Salembier, 2011). For example, previous MTSS-related research found leaders needed to make sure to include a variety of educational professionals to drive the implementation efforts including classroom teachers, specialists, special education teachers, and parents (e.g., Bean & Lillenstein, 2012; Dougherty Stahl et al., 2013). McIntosh and Goodman (2016) noted it was vital for school leadership teams to embed their efforts in a systemic and comprehensive school improvement process. Therefore, the members of the school leadership team should come from a representative group of individuals who work across a wide variety of grade-level teams and teach diverse content areas. According to Kezar (2009), these collaborative endeavors tend to maximize student success.

Additionally, this study also supported and extended previous research that found when school leadership teams clearly communicated, student achievement tended to improve. Including many voices in the scale-up efforts seemed to facilitate the comprehensive buy-in of all staff (e.g., Bean & Lillenstein, 2012; Dougherty Stahl et al., 2013; McIntosh & Goodman, 2016). Similarly, Blasé and Blasé (2000) explained effective instructional leaders have the professional and interpersonal skills to make suggestions, provide feedback, gather opinions, collaborate, facilitate professional development opportunities, and provide praise. Results of this study indicated that creating effective and collaborative leadership teams is a non-negotiable element when MTSS initiatives are being brought to scale. Without powerful leadership teams, schools that struggle to have a positive impact on student learning will continue to fight to have a large and meaningful impact on student learning. This study found a positive correlation exists between perceptions of MTSS implementation, leadership, and student reading outcomes. In sum, the findings of this study supported previous educational leadership research that found in schools with leaders who (a) stayed focused on student learning, (b) facilitated collaborative endeavors, (c) provided technical assistance and appropriate professional development, and (d) guided and participated in the data-based decision-making process, student learning benefited (Kezar, 2009; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Shepherd & Salembier, 2010).

Evidence-Based Practices: Doing What Works

The CDE (2016) defined evidence-based practices as the methods of instruction, intervention, and assessment that have been proven effective in the body of research and resulted in improved outcomes for students. Previous research demonstrated when teachers were able to establish a safe and welcoming environment, keep students engaged in learning, and make the most of their instructional time using strategies that addressed both academics and behavior, students tended to be more successful (McIntosh & Goodman, 2016). Using recommendations of various researchers (e.g., Bradshaw, Waasdorp, & Leaf, 2012; Horner, Sugai, & Anderson, 2010) who shared that evidence-based practices implemented by typical educators have positive, meaningful effects on problem behavior and academic achievement, the survey items forced to load onto the evidence-based practice latent factor included concepts associated with (a) appropriate and just-in time professional development, (b) research-based curricular tools and instructional strategies, and (c) classroom and behavioral management strategies. The results of this study indicated when teachers reported higher levels of access to professional development, evidence-based instructional resources, and felt they had the required skills to effectively manage student behavior, their students tended to have higher student reading outcomes. This study supported and extended the work of various researchers who confirmed the importance of providing teachers with professional development and technical support (e.g., Menzies et al., 2008; Rodriguez & Denti, 2011).

Additionally, previous research demonstrated that when students who struggle while learning to read are provided with evidence-based, differentiated instruction designed to meet their individual needs, they tended to improve (e.g., Case et al., 2010; Smolkowski & Cummings, 2015). This study also supported this body of research, finding when educators reported they used a variety of evidence-based practices in their instructional routines, student reading outcomes tended to improve. In sum, successful scale-up efforts of multi-tiered models that have the potential to positively impact student reading outcomes must provide opportunities for teachers to receive (a) timely and appropriate professional development, (b) technical assistance and (c) instructional support. In the same way teachers provide their students with time and instructional support to learn a concept before mastery is the expectation, educational policy makers and federal legislators must also recognize that classroom teachers need both coaching support and time to master the concepts associated with multi-tiered systems (e.g., Al Otaiba et al., 2016; Deno et al., 2009).

Data-Driven Instruction

In a multi-tiered model, data are used to help drive the instructional planning process, evaluate the effectiveness of curricular tools and programs, and guide overall school improvement efforts. In other words, gathering, analyzing, and using data provide teachers with the information they need to advance learning (Al Otaiba & Fuchs, 2006; Deno et al, 2009; Wanzek & Vaughn, 2007). Results of this study supported the findings of previous researchers who demonstrated when teachers were able to gather, interpret, and use student-level data during the instructional decision-making process, student learning accelerated (e.g., Al Otaiba & Fuchs, 2006; Linan-Thompson et al., 2007).

Multi-tiered system of supports (MTSS) is fundamentally a data-driven initiative. Multiple sources of data are used to monitor fidelity of implementation, student proficiency, and progress and are used during the instructional planning process and during system-level scale-up and implementation efforts (McIntosh & Goodman, 2016). To maximize the reading outcomes of young students, educators must be able to access valid and reliable student-level data to both identify students at risk and intervene early (e.g., Wanzek & Vaughn, 2007). Results of this study indicated when educators reported higher levels of awareness of the types of data they had access to and felt they knew how interpret the data in a meaningful way, student reading outcomes tended to improve. Additionally, when teachers reported higher levels of awareness of how student-level data were gathered, analyzed, and used during the instructional planning process, student reading outcomes also tended to improve.

The data-based problem solving items asked participants to broadly quantify how data were used to guide, refine, and inform the instructional planning process and meet the needs of all students. Generally, a variety of problem-solving processes teams could be utilized to improve student learning outcomes that range from simple to complex. For example, a six-step process endorsed by McIntosh and Goodman (2016) had its origins in (a) the fields of school psychology (Deno, 1995); (b) the team-initiated problem-solving process (TIPS; Newton, Horner, Algozzine, Todd, & Algozzine, 2012); and (c) the outcome-driven model (Good, Gruba, & Kaminski, 2002). Specifically, this six-step process had teams collaboratively (a) identify a problem, (b) gather and analyze data to determine why the problem might be occurring, (c) create a solution, (d) set goals, (e) implement the plan, and (f) evaluate how well the plan worked (McIntosh & Goodman, 2016). Whether using the above six-step problem solving model or an alternative, results of this study indicated when teachers reported they used a data-based problem solving process to meet student needs in their schools, student reading outcomes tended to improve.

Families, Schools, Teachers, and Communities: Working Together

Family, school, and community partnering (FSCP), according to the Colorado Department of Education (2016), describes what happens when families, school professionals, and community members actively communicate and collaborate to improve student learning. This study found a moderate and statistically significant correlation among increased perceptions of parental involvement, the MTSS framework, and student reading outcomes. This finding was supported by previous researchers who shared that high levels of parental involvement in schools tended to positively impact student reading achievement (Senechal & Young, 2008).

The rationale for family, school, and community partnering was derived from over four decades of research that demonstrated how partnerships worked to improve student learning outcomes (Christenson & Reschly, 2010). When schools and families intentionally partnered in an ongoing, sustainable, and intentional manner, student learning tended to improve. The positive effects of partnering were noted for all the individuals involved and included (a) increased student achievement, (b) increased family engagement, (c) higher levels of support for schools at the community level, and (d) increased levels of teacher morale and performance (Eagle, Dowd-Eagle, & Sheridan, 2008).

The MTSS framework is a recent innovation of the 21st century educational reform movement that combines increasingly intensive student-level interventions of the RTI model with the school-wide focus of the SW-PBIS model. Experts hope the combination of the two models into a single, cohesive framework will lead to increased learning outcomes for all students regardless of ability level (e.g., Sugai & Horner, 2009). Based on a thorough review of the empirical research literature, previous study findings demonstrated the components of an MTSS models that might positively correlate with student reading outcomes include (a) concepts associated with collaboration between leaders, teachers, and families; (b) gathering and use of student-level data during the instructional decision-making process; (c) use of instructional practices and curricular tools supported by empirical research; and (d) a system of differentiated instructional supports that provide students with the help they need to proficiently read.

These components are similar to those endorsed by the National Center on Response to Intervention (NCRTI, 2012) of the American Institutes of Research. According to NCRTI, MTSS incorporates assessment and intervention within a multi-level prevention system to increase student achievement and reduce behavior difficulties. With MTSS, schools use data to pinpoint students at risk, monitor progress, provide increasingly intensive instructional supports using evidence-based strategies, and identify students who might benefit from the services provided by an IEP. The four essential components the NCRTI's MTSS model are (a) screening, (b) progress monitoring, (c) multi-level prevention system, and (d) data-based decision-making. The NCRTI also incorporates culturally responsive, evidence-based practices that combine with the four key components to improve the learning outcomes for students (NCRTI, 2012).

Currently, educational leaders from 21 states around America have individually developed their own MTSS models and are collaborating with schools and districts to bring those models to scale. While each of these states identified a variety of similar components, individual state-level models tend to contrast with each other more than they compare. The Colorado components (CDE, 2016) address the four key components of the NCRTI model. Specifically, Colorado's DBPS and TDSL components are closely aligned with the NCRTI (2012) data-based decision-making component because within schools, teams gather and analyze student-level data to make decisions about instruction and identify students who might benefit from additional layers of instructional support. Additionally, Colorado's LCS and EBP components are aligned with the NCRTI multi-level prevention system used by school-level teams to provide students with increasingly intensive levels of support where teachers use evidence-based practices to meet the needs of diverse groups of students.

One of the Colorado components not included within the NCRTI model is associated with family, school, and community partnerships (CDE, 2016). However, other organizations have recognized that student learning outcomes are improved when schools and families collaborate. For example, one of the essential components of a model endorsed by the RTI Action Network (NCRTI, 2012) model includes a component that encourages schools to partner with families to engender significant and meaningful change. Therefore, the Colorado MTSS model both directly and indirectly addressed all of the essential elements leaders at the national-level determined had a meaningful impact on students and their learning. When viewed comprehensively, the results of this study supported the idea that a system like the MTSS framework had the potential to positively impact the reading outcomes of elementary students.

In conclusion, the world we live within is becoming increasingly more diverse and a majority of public schools are serving students from a wide range of cultural and socioeconomic backgrounds. According to Bruce (2008), many students live in homes where money is in short supply and parents need to have more than one job to meet monthly financial obligations. Previous research that examined how poverty impacts students shared a variety of troubling findings. For example, children who live in households with low annual incomes were more likely to experience health problems, have learning disabilities, and be diagnosed with developmental delays than their middle-class peers (e.g., Brooks-Gunn & Duncan, 1998; Pellino, 2007). Additionally, many students from poverty-stricken families did not have the disposable income needed to purchase books nor did they benefit from a large amount of quality time with their parents (Rothstein, 2008). This lack of parental and educational access often resulted in deficits in basic academic skills and cognitive abilities that could have provided a strong foundation for future learning (Bruce, 2008; Butler, 2006; Hampden-Thompson & Johnson, 2006; Nelson, 2006; Pellino, 2007; Rothstein, 2008). Children from homes with low annual incomes were also more likely to suffer from emotional and behavioral problems that negatively impacted their learning than children whose families had a higher socio-economic status (Bruce, 2008). The evidence is clear--poverty negatively impacts student learning.

Fortunately, a significant body of research provides specific strategies educators can employ to offset the effects of poverty on student learning. For example, Clewell and

Campbell (2007) found when high poverty schools were headed by collaborative leaders, students who were impacted by poverty tended to achieve more. Balfanz (2006) concluded students impacted by poverty tended to learn more when they learned in small groups where a range of evidence-based resources were employed. Research also demonstrated that economically disadvantaged students learned more when their teachers worked together and ensured their students received the interventions and instructional supports they needed to grow and achieve (Balfanz, 2006; Kannapel & Clements, 2005). Research also demonstrated that teachers were able to meet the needs of high-poverty students more effectively when they (a) created safe learning environments, (b) provided students with emotional support and encouragement, and (c) communicated and partnered with parents and families (Field, Kuczera, & Pont, 2008). Similarly, when schools that served students impacted by poverty created resource teams to address the factors that might have hindered students' academic progress, their students tended to learn more (Butler, 2006). The body of research also confirmed when teaches used differentiated and flexible pedagogical practices to meet student needs, learning also tended to improve (Educational Research Service, 2001). Finally, schools that were successful at offsetting the effect of poverty on student learning tended to assess student learning on a regular basis (Center for Public Education, 2005).

As the reader will note, most of these ideas are specifically addressed as one of the essential components of the MTSS framework. While previous research demonstrated that individually these factors tended to accelerate the learning of students impacted by poverty, this current study demonstrated when schools combined them into a single, systemic, school-wide framework, student reading achievement accelerated.

Therefore, it can be confidently stated that MTSS has the potential to counteract an important portion of the impact poverty has on the reading outcomes of students who struggle while learning to read. It is an effective system that can be used by educators to have a meaningful and long-term impact on their students, their communities, and the nation at large.

Limitations

This study had four main limitations associated with (a) the novelty of the survey instrument, (b) the sample, (c) topics associated with missing data, and (d) limitations associated with SEM. The first limitation centered on the novelty of the MTSS-IPS as a survey instrument. As noted in previous chapters, the MTSS-IPS is a recently developed survey instrument that was distributed to teachers in western Nebraska and throughout Colorado; only data provided by the Colorado sample of participants were used in the study. While the research team took every precaution to ensure the instrument had high levels of construct and content validity, this study used data generated from the first wide-scale distribution of the instrument and further partitioned the data to analyze information provided by the Colorado sample. While statistical analysis confirmed the MTSS-IPS instrument provided valid and reliable data for this specific group of individuals, the results might be different if used to gather perceptions of MTSS implementation in different states and with different groups of educators. Second, MTSS-IPS data used for this study were generated by a variety of educators and administrators serving elementary students throughout the state of Colorado and focused on gathering perception of MTSS implementation in elementary settings. If the MTSS-IPS was used to gather implementation perception data with individuals serving students in different grades, the findings might be different. Similarly, because individuals with proxy roles were invited to participate, their perceptions of MTSS implementation might have impacted the overall findings. Therefore, the generalizability of this study's findings might be limited. Third, the survey data had larger-than-average levels of non-random missing data. Survey participants who completed the survey might be significantly different from the individuals who started the survey and dropped out before completing all the questions. On average, participants tended to drop out at the end of each page of the electronic survey; thus, the missing data were most likely not missing completely at random. Comparing the responses of target and proxy individuals based on their demographic characteristics provided evidence that the participants did not differ significantly from each other or from the population at large; however, non-response bias might still be a possibility. The final limitation of this study centered around issues associated with SEM methodology. In certain circumstances, standard errors and estimates of fit might not have been correctly estimated and might have increased because a portion of the latent exogenous variables was highly correlated. Secondly, the processes used throughout this study were cognitively challenging and errors might have been unintentionally made that impacted the results. Finally, as noted by Tomarken and Waller (2005), structural models are simply rough estimates of reality. While one proposed model might provide ideal levels of fit, an infinite number of non-identified models could also provide the same level of fit. Therefore, discussion surrounding the results of this study should be interpreted with caution.

Recommendations for Future Research

Recommendations for future MTSS-related research, given the novelty of the initiative, are numerous. First, this study specifically investigated how perceptions of implementation correlated with 2014 third grade reading outcomes and found increased perceptions of implementation were predictive of increased student reading scores. Future research should examine how perception of implementation correlates with student learning outcomes (a) in different grades, (b) using different reading measures, and (c) with different groups of educators serving students in different settings (e.g., public, private, and charter schools). Additionally, future research could also investigate how perceptions of implementation compared in districts with larger and smaller student enrollment numbers and in schools with higher and lower levels of student achievement. Finally, a unique opportunity exists to partner with various state-level departments of education to obtain student reading data from schools and/or districts with very small student counts within individual grades because, as noted, this examination of MTSS implementation perception did not include TCAP reading scores of participants who worked in Colorado schools with less than 16 students. While the information these individuals provided were included in the series of CFAs conducted, they were excluded from the SEM analysis because they worked in very small schools where student reading scores were not publicly reported.

REFERENCES

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? *Journal of Learning Disabilities*, *39*(5), 414–431.

Al Otaiba, S., Folsom, J. S., Wanzek, J., Greulich, L., Waesche, J., Schatschneider, C., & Connor, C. M. (2016). Professional development to differentiate kindergarten tier 1 instruction: Can already effective teachers improve student outcomes by differentiating tier 1 instruction? *Reading & Writing Quarterly, 32*(5), 454–476.

Al Otaiba, S., Kim, Y., Wanzek, J., Petscher, Y., & Wagner, R. K. (2014). Long-term effects of first-grade multitier intervention. *Journal of Research on Educational Effectiveness*, *7*(3), 250–267.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545-557.

Bailey, T. R. (2017, March 23). Re: (RtI) Signing on to our response to the Johns et al. paper [Electronic mailing list message]. Retrieved from rti-bounces@educ. umass.edu

Baker, S. K., Smolkowski, K., Chaparro, E. A., & Fien, H. (2015). Using regression discontinuity to test the impact of a tier 2 reading intervention in first grade. *Journal of Research on Educational Effectiveness, 8*(2), 218–244.

Balfanz, R. (2006, September). *Why are achievement gains so difficult to realize in high poverty middle grade schools? What can be done about it?* Paper presentation at the National Policy Symposium on Middle-Level Education: Where do we go from here? Washington, DC. Retrieved from htttp://www.all4ed.org/files/ Balfanz_AllianceED.ppt

Balu, R., Zhu, P., Doolittle, F., Schiller, E., & Jenkins, J. (2015). *Evaluation of response to intervention practices for elementary school reading*. Retrieved from https://ies.ed.gov/ncee/pubs/20164000/

Barrio, B. L., & Combes, B. H. (2015). General education pre-service teachers' levels of concern on response to intervention (RTI) implementation. *Teacher Education & Special Education, 38*(2), 121–137.

Bean, R., & Lillenstein, J. (2012). Response to intervention and the changing roles of schoolwide personnel. *Reading Teacher, 65*(7), 491–501.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186-203.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238.

Bentler, P. M., & Bonnett (1980). Significant tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Blasé, J., & Blasé, J. (2000). Effective instructional leadership: Teachers' perspectives on how principals promote teaching and learning in schools. *Journal of Educational Administration, 38*(2), 130-141.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley-Interscience Publications.

Boscardin, C. K., Muthen, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100*(1), 192–208.

Bost, L., & Riccomini, P. (2006). Effective instruction: An inconspicuous strategy for dropout prevention. *Remedial and Special Education, 27*(5), 301–311.

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly, 23*(4), 462–473. http://doi.org/10.1037/a0012883

Bradshaw, C. P., Reinke, W. M., Brown, L. D., Bevans, K. B., & Leaf, P. J. (2008). Implementation of school-wide positive behavioral interventions and supports (PBIS) in elementary schools: Observations from a randomized trial. *Education and Treatment of Children, 31*(1), 1–26. http://doi.org/10.1353/etc.0.0025

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*(5), e1136-e1145.

Brantlinger, E., Jimenez, R., Klinger, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children, 71*(2), 195-207.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children, 7*(2), 55-71. Retrieved from www.futureofchildren.org/usr_doc/vol7no2ART4.pdf

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions, 154,* 136-153.

Bruce, C. A. (2008). Countering the effects of poverty on students. *National Association of Elementary School Principals Diverse Learning Communities Today, 1*(1), 1-2.

Byrne, D. S. (1998). *Complexity theory and the social sciences: An introduction*. New York: Routledge

Brynner, J. (2008). Illuminating disadvantage. *Education Journal, 110*, 34–35.

Buckingham, J., Wheldall, K., & Beaman, R. (2012). A randomized control trial of a tier-2 small-group intervention (Mimili) for young struggling readers. *Australian Journal of Learning Difficulties, 17*(2), 79–99.

Bursuck, B., & Blanks, B. (2010). Evidence-based early reading practices within a response to intervention system. *Psychology in the Schools, 47*(5), 421–431.

Butler, S. S. (2006). Low-income, rural elders' perceptions of financial security and health care costs. *Journal of Poverty*, *10*(1), 25-43.

Byrd, R. S. (2005). School failure. *Pediatrics in Review*, *26*(7), 227-237.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models.* New York: Springer-Verlag.

Byrne, B. M. (1998). *Structural equation modeling: Basic concepts, application, and programming.* Mahwah, NJ: Lawrence Erlbaum Associated, Inc.

Calhoon, M. B., Al Otaiba, S., Cihak, D., King, A., & Avalos, A. (2007). Effects of a peer-mediated program on reading skill acquisition for two-way bilingual first-grade classrooms. *Learning Disability Quarterly, 30*(3), 169–184.

Case, L., Speece, D. L., Silverman, R., Ritchey, K. D., Schatschneider, C., Cooper, D. H., & Jacobs, D. (2010). Validation of a supplemental reading intervention for first-grade children. *Journal of Learning Disabilities, 43*(5), 402–417.

Catts, H., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*(2), 163–176.

Center for Public Education. (2005, August). *Key lessons: High-performing, high-poverty schools.* Retrieved from www.centerforpubliceducation.org/site/ c.kjJXJ5MPIwE/ b.1534271/k.9010/Key_Lessons_ Highperforming_highpoverty_schools.html

Chard, D. J., Stoolmiller, M., Harn, B. A., Wanzek, J., Vaughn, S., Linan-Thompson, S., & Kame'enui, E. J. (2008). Predicting reading success in a multilevel schoolwide reading model: A retrospective analysis. *Journal of Learning Disabilities, 41*(2), 174–188.

Christenson, S. L., & Reschly, A. L. (2010). *Handbook of school–family partnerships.* New York, NY: Routledge.

Clemens, N. H., Hilt-Panahon, A., Shapiro, E. S., & Yoon, M. (2012). Tracing student responsiveness to intervention with early literacy skills indicators: Do they reflect growth toward text reading outcomes? *Reading Psychology, 33*(1/2), 47–77.

Clewell, B. C., & Campbell, P. B. (2007). *Good schools in poor neighborhoods: Defying demographics, achieving success.* Baltimore, MD: Urban Institute Press.

Colorado Department of Education. (2014). *About CSAP/TCAP*. Retrieved from https://www.cde.state.co.us/assessment/coassess-about

Colorado Department of Education. (2016, June). *Colorado multi-tiered system of supports (CO-MTSS)*. Retrieved from https://www.cde.state.co.us/mtss/mtssessentialcomponentsdefinitionsjune2016

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., …Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*(3), 204–216.

Comrey, A. L. & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage Publications.

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

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934–945. http://doi.org/10.1037/0012-1649.33.6.934

Deno, S. L. (1995). School psychologist as problem solver. *Best Practices in School Psychology, 3*, 471-484.

Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., & Stachel, N. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools, 46*(1), 44–55.

Denton, C. A., Cirino, P. T., Barth, A. E., Romain, M., Vaughn, S., Wexler, J., …Fletcher, J. M. (2011). An experimental study of scheduling and duration of tier 2 first-grade reading intervention. *Journal of Research on Educational Effectiveness*, *4*(3), 208–230.

Denton, C. A., Fletcher, J. M., & Anthony, J. L. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities, 39*(5), 447–466.

Denton, C. A., Kethy, C., Nimon, K., Kurz, T. B., Mathes, P. G., Minyi, S., & Swanson, E. A. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children, 76*(4), 394–416.

Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A. E., Vaughn, S., & Francis, D. J. (2013). Effects of tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology, 105*(3), 633–648.

Dougherty Stahl, K. A., Keane, A. E., & Simic, O. (2013). Translating policy to practice: Initiating RTI in urban schools. *Urban Education, 48*(3), 350–379.

Eagle, J. W., Dowd-Eagle, S. E., & Sheridan, S. M. (2008). Best practices in school-community partnerships. *Best Practices in School Psychology 5*(3), 953-967.

Educational Research Service. (2001). *What can schools do to reduce the achievement gap? On the Same Page series.* Retrieved from www.ers.org/otsp/otsp3.html

Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95–113. http://doi.org/10.2307/2112677

Field, S., Kuczera, M., & Pont, B. (2008). *No more failures: Ten steps to equity in education.* Retrieved from https://www.oecd.org/education/school/45179151.pdf

Fiester, L. (2010). *Early warning! Why reading by the end of third grade matters.* Retrieved from http://www.aecf.org/m/resourcedoc/AECF-Early_Warning_Full_Report-2010.pdf

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37. http://doi.org/10.1111/j.1750-8606.2008.00072.x

Fowler, F. J. (2008). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making secondary intervention work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading & Writing, 21*(4), 413–436.

Fuchs, D., Compton, D., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347.

Fuchs, L. S., & Fuchs, D. (2006). Implementing responsiveness-to-intervention to identify learning disabilities. *Perspectives on Dyslexia, 32*(1), 39–43.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195–203.

Garbacz, S. A., McIntosh, K., Eagle, J. W., Dowd-Eagle, S. E., Hirano, K. A., & Ruppert, T. (2016). Family engagement within schoolwide positive behavioral interventions and supports. *Preventing School Failure: Alternative Education for Children and Youth, 60*(1), 60-69.

Gehsmann, K. M. (2008). Response to intervention: What is it and how do schools and districts prepare for it? *Michigan Reading Journal*, *40*(3), 22–30.

Gelzheiser, L. M., Scanlon, D., Vellutino, F., Hallgren-Flynn, L., & Schatschneider, C. (2011). Effects of the interactive strategies approach—extended: A responsive and comprehensive intervention for intermediate-grade struggling readers. *The Elementary School Journal, 112*(2), 280-306.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention (RtI) and multi-tier intervention in the primary grades. A practice guide* (NCEE2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.

Gil, L., & Woodruff, D. (2011). Connecting research to promising practice for English Learners. *School Administrator, 68*(5), 10–16.

Gilbert, J., Compton, D., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, *48*(2), 135–154.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022440506000562

Good, R. H., & Kaminski, R. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., & Kaminski, R. A. (Eds.). (2007). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). *Best practices in using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model*. Retrieved from http://dibels.uoregon.edu

Graves, A. W., Duesbery, L., Pyle, N. B., Brandon, R. R., & McIntosh, A. S. (2011). Two studies of tier II literacy development. *Elementary School Journal*, *111*(4), 641–661.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science, 23*(4), 519-529.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: Wiley & Sons, Inc.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72*(2), 167-189.

Hagans, K. S. (2008). A response-to-intervention approach to decreasing early literacy differences in first graders from different socioeconomic backgrounds. *Assessment for Effective Intervention, 34*(1), 35–42.

Hampden-Thompson, G., & Johnson, J. S. (2006). *Variation in the relationships between non-school factors and student achievement on international assessments.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Harn, B. A., Chard, D. J., Biancarosa, G., & Kame'enui, E. (2011). Coordinating instructional supports to accelerate at-risk first-grade readers' performance. *Elementary School Journal, 112*(2), 332–355.

Harn, B. A., Chard, D. J., & Kame-enui, E. J. (2011). Meeting societies' increased expectations through responsive instruction: The power and potential of system-wide approaches. *Preventing School Failure, 55*(4), 232–239.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53-60. Retrieved from http://www.ejbrm.com

Hoppey, D. (2013). Linking action research to response to intervention (RtI): The strategy implementation project. *Networks: An Online Journal for Teacher Research, 15*(1), 1–10.

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8), 1.

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

Hudson, T. M. (2013). *Responding to the need for research in RTI: A map of the current implementation and assessment practices from the perspective of district and building administrators.* Retrieved from ProQuest LLC (UMI Accession No. 3579743).

Hughes, C. A., & Dexter, D. D. (2011). Response to intervention: A research-based summary. *Theory into Practice, 50*(1), 4–11.

Hurst, S. (2014, January 6). *Re: What is the difference between RTI and MTSS?* [Web log comment]. Retrieved from http://www.readinghorizons.com/blog/what-is-the-difference-between-rti-and-mtss

International Business Machines Corporation. (2016). *IBM SPSS Statistics for Windows* (Version 24.0) [Computer software]. Armonk, NY: IBM Corp.

Jacobs, J., Gregory, A., Hoppey, D., & Yendol-Hoppey, D. (2009). Data literacy: Understanding teachers' data use in a context of accountability and response to intervention. *Action in Teacher Education, 31*(3), 41–55.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582–600.

Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S. K., & Tilly, W. D. (2013). Responsiveness to intervention in reading: Architecture and practices. *Learning Disability Quarterly, 36*(1), 36–46.

Johnston, P. (2010). An instructional frame for RTI. *Reading Teacher, 63*(7), 602–604.

Joreskog, K. & Sorbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International Inc.

Josephson, A. (2015, December). *Why your retirement age matters*. Retrieved from https://smartasset.com/retirement/why-your-retirement-age-matters

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447. http://doi.org/10.1037/0022-0663.80.4.437

Kame'enui, E. J. (2007). A new paradigm: Responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 6–7.

Kansas Multi-Tiered System of Supports. (2010). *The integration of MTSS and RtI.* Retrieved from http://www.kansasmtss.org/pdf/briefs/The_Integration_of_MTSS_and_RtI.pdf

Kannapel, P. J., & Clements, S. K. (2005). *Inside the black box of high-performing high-poverty schools.* Retrieved from http://www.cdl.org/resource-library/pdf/FordReportJE/pdf.

Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage Publications.

Keith, T. Z. (1999). Structural equation modeling in school psychology. *The Handbook of School Psychology, 3*(1), 78-107.

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., …Velez, E. D. V. (2015). *The condition of education 2015*. NCES 2015-144. National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Kerins, M. R., Trotter, D., & Schoenbrodt, L. (2010). Effects of a tier 2 intervention on literacy measures: Lessons learned. *Child Language Teaching & Therapy, 26*(3), 287–302.

Kezar, A. (2009). Change in higher education: Not enough, or too much? *Change: The Magazine of Higher Learning, 41*(6), 18-23.

Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741–756. http://doi.org/10.1037/a0019237

Klem, L. (2000). Structural equation modeling. In L. G. Grim & P. R. Yamold (Eds.), *Reading and understanding more multivariate statistics* (pp. 227-559). Washington DC: American Psychological Association.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: The Guilford Press.

Lane, K. L., Menzies, H. M., Ennis, R. P., & Bezdek, J. (2013). School-wide systems to promote positive behaviors and facilitate instruction. *Journal of Curriculum and Instruction, 7*(1), 6–31. http://doi.org/10.3776/joci.2013.v7n1p6-31

Lane, K. L., Oakes, W. P., & Menzies, H. M. (2014). Comprehensive, integrated, three-tiered models of prevention: Why does my school-and district-need an integrated approach to meet students' academic, behavioral, and social needs? *Preventing School Failure: Alternative Education for Children and Youth, 58*(3), 121–128. http://doi.org/10.1080/1045988X.2014.893977

LaVigna, G. W., & Willis, T. J. (2012). The efficacy of positive behavioural support with the most challenging behaviour: The evidence and its implications. *Journal of Intellectual & Developmental Disability, 37*(3), 185–95. http://doi.org/10.3109/13668250.2012.696597

Lei, P., & Wu, Q. (2007). Introduction to structural equation modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more about multivariate statistics* (pp. 227-260). Washington, DC: American Psychological Association.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Learning from leadership: A review of the literature.* Minneapolis, MN: University of Minnesota, Center for Applied Research and Educational Improvement.

Linan-Thompson, S., Cirino, P. T., & Vaughn, S. (2007). Determining English language learners' response to intervention: Questions and some answers. *Learning Disability Quarterly, 30*(3), 185–195.

Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. T. (2006). The Response to Intervention of English Language Learners at risk for reading problems. *Journal of Learning Disabilities, 39*(5), 390–398.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130.

MacGintie, W., & MacGintie, R. (2006). *Gates-MacGinitie reading tests* (4th ed.). Iowa City, IA: Houghton Mifflin.

Mathes, P. G., Denton, C., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*(2), 148–182. http://doi.org/10.1598/RRQ.40.2.2

McAlenney, A. L., & Coyne, M. D. (2015). Addressing false positives in early reading assessment using intervention response data. *Learning Disability Quarterly*, *38*(1), 53–65.

McCombes-Tolis, J., & Spear-Swerling, L. (2011). The preparation of preservice elementary educators in understanding and applying the terms, concepts, and practices associated with response to intervention in early reading contexts. *Journal of School Leadership, 21*(3), 360–389.

McIntosh, K., & Goodman, S. (2016). *Integrated multi-tiered systems of support: Blending RTI and PBIS.* New York: Guilford Publications.

McIntosh, K., Horner, R. H., Chard, D. J., Dickey, C. R., & Braun, D. H. (2008). Reading skills and function of problem behavior in typical school settings. *The Journal of Special Education*, *42*(3), 131–147. http://doi.org/10.1177/0022466907313253

McIntosh, K., Reinke, W. M., & Herman, K. C. (2009). Schoolwide analysis of data for social behavior problems. In G. G. Peacock, R. A. Erwin, E. J. Daly, & K. J. Merrell (Eds.), *Practical handbook of school psychology: Effective practices for the 21st century* (pp. 135-156). New York: Guilford Publications.

Menzies, H. M., Mahdavi, J. N., & Lewis, J. L. (2008). Early intervention in reading: From research to practice. *Remedial & Special Education*, *29*(2), 67–77.

Mesmer, E. M., & Mesmer, H. A. E. (2008). Response to intervention (RTI): What teachers of reading need to know. *Reading Teacher*, *62*(4), 280–290.

Miller, A. C., Fuchs, D., Fuchs, L. S., Compton, D., Kearns, D., Zhang, W., …Kirchner, D. P. (2014). Behavioral attention: A longitudinal study of whether and how it influences the development of word reading and reading comprehension among at-risk readers. *Journal of Research on Educational Effectiveness, 7*(3), 232–249.

Millsap, R. E. (2007). Structural equation modeling made difficult. *Personality and Individual Differences, 42*(5), 875-881.

Moats, L. (2009). Knowledge foundations for teaching reading and spelling. *Reading & Writing*, *22*(4), 379–399.

Mokhtari, K., Neel, J. L., Kaiser, F., & Hong-Hai, L. (2015). Assessing the promise of a supplemental reading intervention for at-risk first grade students in a public school setting. *International Electronic Journal of Elementary Education, 7*(3), 281–300.

Muthen, L. K., & Muthen, B. O. (2015). *MPlus* (Version 7.4) [Computer software]. Los Angeles, CA: Author.

National Center on Response to Intervention. (2012) *Essential components of RTI.* Retrieved from http://www.rti4success.org/

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from https://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academies Press.

National School Climate Center. (2015). *School climate*. Retrieved from http://www.schoolclimate.org/climate/

NCS Pearson, Inc. (2014). *AIMSweb reading*. New York, NY: Pearson.

Nelson, A. (2006). *Overcoming the income gap*. Retrieved from http://www.ascd.org/publications/newsletters/infobrief/fall06/num47/toc.aspx

Newton, J. S., Horner, R. H., Algozzine, B., Todd, A. W., & Algozzine, K. (2012). A randomized wait-list controlled analysis of the implementation integrity of team-initiated problem solving processes. *Journal of School Psychology, 50*(4), 421-441.

No Child Left Behind Act of 2001, PL 107-110, 115 Stat. 1425, 20 U. S. C. §§ 6301 *et seq.* (2001).

Northwest Evaluation Association. (2017, February 22). *Measure student progress with MAP*. Retrieved from https://www.nwea.org/assessments/map/

Oakes, W. P., Lane, K. L., & Germer, K. A. (2014). Developing the capacity to implement tier 2 and tier 3 supports: How do we support our faculty and staff in preparing for sustainability? *Preventing School Failure: Alternative Education for Children and Youth, 58*(3), 183–190. http://doi.org/10.1080/1045988X.2014.895575

Oakes, W. P., Mathur, S. R., & Lynne, K. (2010). Reading interventions for students with challenging behavior: A focus on fluency. *Behavioral Disorders, 35*(2), 120–139.

O'Connor, R. E., Bocian, K., Beebe-Frankenberger, M., & Linklater, D. L. (2010). Responsiveness of students with language difficulties to early intervention in reading. *Journal of Special Education, 43*(4), 220–235.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B, & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*, 137-148.

Orosco, M., & Klinger, J. (2010). One school's implementation of RTI with English language learners: Referring into RTI. *Journal of Learning Disabilities, 43*(3), 269–288.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3*, 96-146.

Pellino, K. M. (2007). *The effects of poverty on teaching and learning*. Retrieved from www.technology.com/tutorials/teaching/poverty/print.html

Peters, C. L. O., & Enders, C. (2002). A primer for the estimation of structural equation models in the presence of missing data. *Journal of Targeting, Measurement, and Analysis for Marketing, 11*, 81-95.

Pierce, C. D., Klopfenstein, K., & Mathis, J. (n.d.). *The multi-tiered system of supports implementation perception survey*. Unpublished manuscript.

Planty, M., Hussar, W. J., & Snyder, T. D. (2008). *Condition of education 2008* (NCES 2008-031). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Popham, W. J. (2009). Assessment literacy for teachers: Fadish or fundamental? *Theory into Practice, 48*(1), 4-11. doi:10.1080//00405840802577536.

Pyle, A., Wade-Woolley, L., & Hutchinson, N. L. (2011). Just listen to us: The role of teacher empowerment in the implementation of responsiveness to intervention. *Alberta Journal of Educational Research, 57*(3), 258–272.

Regan, K., Berkeley, S., Hughes, M., & Brady, K. (2015). Understanding practitioner perceptions of responsiveness to intervention. *Learning Disability Quarterly, 38*(4), 234–247.

Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavior Interventions, 15*(1), 39–50. http://doi.org/10.1177/1098300712459079

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children’s mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, *26*(1), 1–13. http://doi.org/10.1037/a0022714

Remler, D. K., & Van Ryzin, G. G. (2015). *Research methods in practice: Strategies for description and causation* (2nd ed.). Los Angeles, CA: Sage Publications.

Ritchey, K. K., Silverman, R. D., Montanaro, E. A., Speece, D. L., & Schatschneider, C. (2012). Effects of a tier 2 supplemental reading intervention for at-risk fourth-grade students. *Exceptional Children*, *78*(3), 318–334.

Rodriguez, C., & Denti, L. (2011). Improving reading for second grade English language learners using an RTI approach. *California Reader, 44*(4), 12–18.

Rose, J. (2006). *Independent review of early reading: Final report*. London: Department for Education and Science.

Rothstein, R. (2008). Whose problem is poverty*? Educational Leadership*, *65*(7), 8-13.

Samuels, C. A. (2016, December 13). *RTI expands, encounters growing pains*. Retrieved from http://www.edweek.org/ew/articles/2016/12/14/rti-expands-encounters-growing-pains.html

Scanlon, D. M. (2013). RTI implementation survey yields both promising and troubling findings. *Reading Today, 31*(3), 6-7.

Scanlon, D. M., Gelzheiser, L. M., Vellutino, F. R., Schatschneider, C., & Sweeney, J. M. (2008). Reducing the incidence of early reading difficulties: Professional development for classroom teachers versus direct interventions for children. *Learning and Individual Differences*, *18*(3), 308-315.

Schatschneider, C., Wagner, R. K., & Crawford, E. C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning & Individual Differences, 18*(3), 308–315.

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

Schuele, C. M., Justice, L. M., Cabell, S. Q., Knighton, K., Kingery, B., & Lee, M. W. (2008). Field-based evaluation of two-tiered instruction for enhancing kindergarten phonological awareness. *Early Education & Development, 19*(5), 726–752.

Senechal, M., & Young, L. (2008). The effect of family literacy interventions on children’s acquisition of reading from kindergarten to grade 3: A meta-analytic review. *Review of Educational Research, 78*(4), 880–907.

Shepherd, K., & Salembier, G. (2010). Leading, learning and literacy: Implementing a response to intervention approach in the Riverside Elementary School. *New England Reading Association Journal, 46*(1), 38–48.

Shepherd, K., & Salembier, G. (2011). Improving schools through a response to intervention approach: A cross-case analysis of three rural schools. *Rural Special Education Quarterly, 30*(3), 3–15.

Sherman, V. J. H. (2016). *Multi-tiered system of supports implementation and scale-up: A case study.* Manuscript submitted for publication.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*(4), 601–617.

Smolkowski, K., & Cummings, K. D. (2015). Evaluation of diagnostic systems: The selection of students at risk of academic difficulties. *Assessment for Effective Intervention, 41*(1), 41-54.

Snow, C., Burns, M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Research Council.

Spear-Swerling, L. S., & Cheesman, E. E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading & Writing, 25*(7), 1691–1723.

Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *Elementary School Journal, 111*(4), 585–607.

Stanovich, K. E. (1986). Matthew effect in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180.

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

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.

Streiner, D. L. (2003). Being inconsistent about inconsistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*(3), 217-222.

Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review, 35*(2), 245–259.

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

Suhr, D. (2008, November). *Step your way through path analysis.* Retrieved from http://lexjansen.com/wuss/2008/pos/pos04.pdf

Swanson, E., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disability Quarterly*, *35*(2), 115–126.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65.

Torgesen, J. K. (2009). The response to intervention instructional model: Some outcomes from a large-scale implementation in reading first schools. *Child Development Perspectives*, *3*(1), 38–40. http://doi.org/10.1111/j.1750-8606.2009.00073.x

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response.* Cambridge, MA: Cambridge University Press.

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

Turnbull, A., Turnbull, R., & Wehmeyer, M.L. (2010). *Exceptional lives: Special education in today's schools.* Upper Saddle River, NJ: Pearson.

Ullman, J. B. (2006). Structural equation modeling: Reviewing the basics and moving forward. *Journal of Personality Assessment, 87*(1), 35-50.

Ulmer, T. (2004). *An investigation of the factors that influence the involvement of school psychologists in transition services* (Doctoral dissertation). Available from ProQuestion Dissertations and Theses database. (UMI No. 3139633)

U.S. Department of Agriculture. (2016, March*). Child nutrition programs: Income eligibility guidelines.* Retrieved from www.federalregister.gov/documents/2016/03/23/2016-06463/child-nutrition-programs-income-eligibility-guidelines.

U.S. Department of Education. (2004). *Individuals with Disabilities Education Act of 2004*. Retrieved from https://www2.ed.gov/policy/speced/guid/idea/idea2004.html

U.S. Department of Education. (2016). *The federal role in education*. Retrieved from https://www2.ed.gov/about/overview/fed/role.html

Utley, C. A., & Obiakor, F. E. (2015). Special issue: Research perspectives on multi-tiered system of support. *Learning Disabilities: A Contemporary Journal, 13*(1), 1–2.

Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75*(2), 165–183.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601–638. http://doi.org/10.1037/0022-0663.88.4.601

Vellutino, F. R., Scanlon, D. M., Zhang, H., & Schatschneider, C. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading & Writing*, *21*(4), 437–480.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541–561.

Wilcox, K. A., Murakami-Ramalho, E., & Urick, A. (2013). Just-in-time pedagogy: Teachers' perspectives on the response to intervention framework. *Journal of Research in Reading*, *36*(1), 75–95.

Yong, A. G., & Pierce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94.

Zirkel, P. A. (2011). RTI confusion in the case law and the legal commentary. *Learning Disability Quarterly*, *34*(4), 242–247. http://doi.org/10.1177/0731948711421760

APPENDIX A

DYNAMIC INDICATOR OF BASIC EARLY LITERACY SKILLS

One commonly used universal screening and progress monitoring tool is DIBELS (Good & Kaminski, 2002). The DIBELS assesses students' initial risk factors and early reading proficiency skills using a variety of measures that include letter naming fluency (LNF), first sound fluency (FSF), phoneme segmentation fluency (PSF), nonsense word fluency (NWF), and oral reading fluency (ORF).

The letter naming fluency measure requires young students in prekindergarten or kindergarten to name as many upper and lower case letters as possible in one minute. While LNF is not linked to an early reading skill because the measure is highly predictive of later reading success, it is included to help identify the lowest 20% of students in a school or district at risk of developing reading difficulties and detect students with scores that fall between the 20^{th} and 40^{th} percentile who might have some risk of developing difficulties in the future (Good & Kaminski, 2002).

The initial sound fluency (ISF) measure is used with young children in pre-school or kindergarten to measure their ability to isolate and orally generate the first sound, or phoneme, in a given word. Benchmark expectations for mid-year kindergarten range between 25 and 35 correct initial sounds; results suggest students who isolate fewer than 10 initial sounds during the mid-year benchmarking period might benefit from more intensive interventions and support (Good & Kaminski, 2002).

Phoneme segmentation fluency (PSF), the third of the DIBLES subscales, is used with students from the middle of kindergarten through the end of first grade and measures students' ability to isolate words with groups of three and four phonemes into their individual segments (Good & Kaminski, 2002). For example, the word "ham" has three phonemes (h/a/m), which students would be expected to individually isolate. Any phonemes that are blended (e.g., h/am, ha/m, or /ham/) are scored as a single sound. The benchmark goal for students is 35 to 45 correct phonemes in one minute at both the end of kindergarten and the beginning of first grade alike. Students who are not able to isolate 10 phonemes at the end of kindergarten and/or the beginning of first grade, according to the developers, might need more intensive supports to meet grade-level benchmark objectives.

Nonsense word fluency (NWF; Good & Kaminski, 2002) is used with students from the middle of kindergarten to the beginning of second grade and assesses students' knowledge with the alphabetic principle. As the name implies, students are shown a list of consonant-vowel-consonant and vowel-consonant nonsense words (e.g., vaj, dit, ab, ot) and asked to read as many of those words (either sound by sound, as whole words, or using a combination of the two techniques) as possible in one minute. The mid-year grade-level benchmarking goal for first grade students is 50 correct letter sounds (CLS) per minute.

Finally, the oral reading fluency (ORF) subtest can be used with first grade students during the midyear benchmarking window through the end of sixth grade. Passages of connected text that vary by grade-level are provided to the student who reads the text out loud for one minute. Individual words that are omitted, incorrectly read, or make a student hesitate for longer than three seconds are counted as errors and subtracted from the total. The number of words correctly read in one minute is the ORF score.

APPENDIX B

CENTER ON RESPONSE TO INTERVENTION SCREENING TOOLS CHART

Screening Tools Chart

As of May 2014

This tools chart reflects the results of the fourth annual review of screening tools by the Center's Technical Review Committee (TRC).

> *The Center defines screening as follows: Screening involves brief assessments that are valid, reliable, and evidence-based. They are conducted with all students or with targeted groups of students to identify those who are at risk of academic failure and, therefore, are likely to need additional or alternative forms of instruction to supplement the conventional general education approach.*

Chart Features

- Across the top of the chart are the standards by which the TRC reviews each tool. When viewing the online version of the chart, click on each standard for a detailed description of how the rating was defined.
- The vendors/developers of the tools have provided implementation information that includes the cost of the tool, what is needed to implement it, the support provided, how the tool is intended to be used, and with whom it should be used. To access this information when viewing the online version of the chart, click on the name of the tool in the "Area" column.
- To view the specific data submitted for Classification Accuracy, Generalizability, Reliability, Validity, and Disaggregated Data for Diverse Populations when viewing the online version of the chart, click the ratings in the chart.
- When viewing the online version of the chart:
  - Every column of the chart can be sorted by clicking the arrows at the tops of the columns.
  - The tools in the chart can be filtered by subject and by grade using the filter tool at the top of the chart. To see all tools again, click "Reset."
  - Tools can be compared by clicking the boxes on the far right of the chart. Select as many tools as you wish to compare and click the "Compare" button. To see all tools again, click "Reset."

The Center on Response to Intervention at American Institutes for Research publishes this chart to assist educators and families in becoming informed consumers who can select screening tools that best meet their individual needs. The Center's Technical Review Committee (TRC) on Screening independently established criteria for evaluating the scientific rigor of screening tools. The TRC rated each submitted tool against these criteria but did not compare it to other tools on the chart. The presence of a particular tool on the chart does not constitute endorsement and should not be viewed as a recommendation from either the TRC on Screening or the Center on Response to Intervention. Please note that all submissions to the TRC review process were voluntary.

Screening Tools Chart

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
A+® LearningLink™- Progress in Math	Mathematics	◐	Moderate Low	◐	◐	—	Group	35-40 minutes	Computer Scored	Yes
Acuity	English Language Arts	◐	Moderate High	◐	●	—	Group	50 minutes	Yes	Yes
	Mathematics	◐	Moderate High	◐	●	—	Group	50 minutes	Yes	Yes
AIMSweb	Mathematics–Curriculum-Based Measurement	◐	Moderate High	◐	○	—	Group	2 minutes	Yes	Yes
	Mathematics Concepts and Applications	◐	Moderate Low	●	○	○	Individual Group	11–13 minutes	Yes	Yes
	Reading–Curriculum-Based Measurement	◐	Moderate High	●	◐	◐	Individual	1–5 minutes	Yes	Yes
	Test of Early Literacy–Letter Naming Fluency	●	Moderate Low	●	◐	—	Individual	2 minutes	Yes	Yes
	Test of Early Numeracy–Missing Number	○	Broad	●	●	—	Individual	2 minutes	Yes	Yes
	Test of Early Numeracy–Number Identification	○	Broad	●	○	—	Individual	2 minutes	Yes	Yes
	Test of Early Numeracy–Oral Counting	○	Moderate Low	○	○	—	Individual	2 minutes	Yes	Yes
	Test of Early Numeracy–Quantity Discrimination	○	Broad	●	●	—	Individual	2 minutes	Yes	Yes
Classworks Universal Screener	Mathematics	◐	Moderate High	●	○	—	Group	30 minutes	Computer Scored	Yes
	Reading	◐	Moderate High	●	○	—	Group	30 minutes	Computer Scored	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
DIBELS 6th Edition	*Letter Naming Fluency	◐	Moderate Low	●	○	—	Individual	2 Minutes	No	Yes
	*Nonsense Word Fluency	●	Moderate Low	●	◐	◐	Individual	2 Minutes	No	Yes
	*Oral Reading Fluency	●	Moderate Low	●	◐	—	Individual	2 Minutes	No	Yes
	*Phoneme Segmentation Fluency	◐	Moderate Low	◐	○	—	Individual	2 Minutes	No	Yes
DIBELS Next	*Daze (DIBELS Maze)	◐	Moderate High	●	○	○	Individual Group	3-6 Minutes	Yes	Yes
	*DORF (DIBELS Oral Reading Fluency)	◐	Moderate High	●	○	◐	Individual	1-2 Minutes	Yes	Yes
	*First Sound Fluency	○	Moderate Low	◐	○	—	Individual	1-3 Minutes	Yes	Yes
	*Nonsense Word Fluency	◐	Moderate High	○	○	○	Individual	1 Minute	Yes	Yes
	*Phoneme Segmentation Fluency	○	Moderate Low	○	○	—	Individual	1-2 Minutes	Yes	Yes
Discovery Education Predictive Assessment	Mathematics	●	Moderate High	●	◐	◐	Group	40 minutes	Yes	Yes
	Reading	●	Moderate High	●	◐	◐	Group	40 minutes	Yes	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
easyCBM	Mathematics	●	Moderate High	◐	●	●	Individual Group	30 minutes	Computer Scored	Yes
	Multiple Choice Reading Comprehension	●	Moderate High	○	○	◐	Individual Group	25–40 minutes	Computer Scored	Yes
	Passage Reading Fluency	●	Moderate High	—	○	◐	Individual	3–4 minute	Yes	Yes
	Vocabulary	●	Moderate High	—	◐	◐	Group	15 minutes	Computer Scored	Yes
EdcheckupStandard Reading Passages	Maze	◐	Moderate High	●	○	●	Group	20 minutes	Yes	Yes
	Oral Reading Fluency	●	Moderate High	●	●	●	Individual	15 minutes	Yes	Yes
Formative Assessment System for Teachers (FAST): Adaptive Math	*aMath	●	Moderate Low	●	◐	—	Individual	10–45 Minutes	Yes	Yes
Formative Assessment System for Teachers (FAST): Adaptive Reading	Reading	●	Moderate Low	●	●	—	Individual Group	6–20 Minutes	Computer Scored	Yes
FAST CBMReading	*English	●	Moderate Low	●	●	●	Individual	1–5 Minutes	Yes	Yes
FAST CBMReading Spanish	*Spanish	◐	Moderate Low	●	○	—	Individual	1–5 Minutes	Yes	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
FAST earlyReading English	*Composite	●	Moderate Low	◐	●	—	Individual	5 Minutes	Yes	Yes
	*Concepts of Print	◐	Moderate Low	○	○	—	Individual	1.5–2.5 Minutes	Yes	Yes
	*Decodable Words	●	Moderate Low	●	○	◐	Individual	1–2 Minutes	Yes	Yes
	*Letter Names	◐	Moderate Low	●	○	◐	Individual	1–1.5 Minutes	Yes	Yes
	*Letter Sounds	●	Moderate Low	●	○	◐	Individual	1–2 Minutes	Yes	Yes
	*Nonsense Words	●	Moderate Low	●	○	◐	Individual	1–2 Minutes	Yes	Yes
	*Onset Sounds	●	Moderate Low	●	○	○	Individual	2–3 Minutes	Yes	Yes
	*Rhyming	●	Moderate Low	●	○	—	Individual	2–3 Minutes	Yes	Yes
	*Sentence Reading	●	Moderate Low	●	●	◐	Individual	1–2 Minutes	Yes	Yes
	*Sight Words (50)	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Sight Words (150)	●	Moderate Low	●	○	◐	Individual	1–2 Minutes	Yes	Yes
	*Word Blending	●	Moderate Low	●	○	○	Individual	1–3 Minutes	Yes	Yes
	*Word Segmenting	●	Moderate Low	●	○	○	Individual	1–3 Minutes	Yes	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
FAST earlyReading Spanish	*Concepts of Print	○	Moderate Low	●	—	—	Individual	1.5–2.5 Minutes	Yes	Yes
	*Decodable Words	◐	Moderate Low	●	○	—	Individual	1–2+ Minutes	Yes	Yes
	*Letter Names	◐	Moderate Low	●	○	—	Individual	1–1.5 Minutes	Yes	Yes
	*Letter Sounds	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Onset Sounds	●	Moderate Low	●	○	—	Individual	2–3 Minutes	Yes	Yes
	*Rhyming	◐	Moderate Low	●	○	—	Individual	2–3 Minutes	Yes	Yes
	*Sentence Reading	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Sight Words (50)	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Sight Words (150)	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Word Blending	◐	Moderate Low	●	○	—	Individual	1–3 Minutes	Yes	Yes
	*Syllables	◐	Moderate Low	●	○	—	Individual	1–2 Minutes	Yes	Yes
	*Word Segmenting	○	Moderate Low	●	○	—	Individual	1–3 Minutes	Yes	Yes
Gates-MacGinitie Reading Tests (GMRT)	Reading	●	Moderate Low	●	○	—	Group	55 minutes	Yes	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency: Administration Format	Efficiency: Administration and Scoring Time	Efficiency: Scoring Key	Efficiency: Benchmarks / Norms
Group Assessment and Diagnostic Evaluation	Group Mathematics Assessment and Diagnostic Evaluation (G-MADE)	◐	Moderate Low	●	◐	—	Individual Group	46–95 minutes	Yes	Yes
	Group Reading Assessment and Diagnostic Evaluation (GRADE)	◐	Moderate Low	●	◐	—	Individual Group	46–95 minutes	Yes	Yes
Iowa Tests of Basic Skills (ITBS)	Mathematics	●	Moderate High	◐	●	—	Group	60 minutes	Yes	Yes
	Reading	●	Moderate High	◐	●	—	Group	55 minutes	Yes	Yes
istation Indicators of Progress	Reading	◐	Moderate Low	●	●	◐	Individual Group	13–21 minutes	Yes	Yes
mCLASS	Mathematics	○	Moderate High	●	○	○	Individual Group	1–12 minutes	Yes	Yes
	Vocabulary Assessment	○	Moderate Low	◐	●	○	Individual	1–2 minutes	Yes	Yes
	**3D–Text Reading and Comprehension	◐	Moderate High	◐	○	●	Individual	5–8 Minutes	Yes	Yes
Measures of Academic Progress (MAP)	Mathematics	●	Moderate High	◐	◐	◐	Individual Group	40 minutes	Computer Scored	Yes
	Reading	●	Moderate High	◐	◐	◐	Individual Group	40 minutes	Computer Scored	Yes
Measures of Academic Progress (MAP) for Primary Grades	Mathematics	●	Moderate High	◐	◐	◐	Individual Group	40 minutes	Computer Scored	Yes
	Reading	●	Moderate High	◐	◐	◐	Individual Group	40 minutes	Computer Scored	Yes

Legend ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

Screening Tools Chart Continued

TOOLS	AREA	Classification Accuracy	Generalizability	Reliability	Validity	Disaggregated Reliability, Validity, and Classification Data for Diverse Populations	Efficiency			
							Administration Format	Administration and Scoring Time	Scoring Key	Benchmarks / Norms
Observation Survey of Early Literacy Achievement	Reading	●	Broad	●	●	●	Individual	15–45 minutes	Yes	Yes
PALS	Early Literacy (Kindergarten)	●	Moderate High	●	◐	●	Individual Group	23–43 minutes	Yes	Yes
	Reading (Grades 1–3)	●	Moderate High	●	●	◐	Individual Group	23–43 minutes	Yes	Yes
Predictive Assessment of Reading	Reading	●	Broad	●	●	●	Individual	16 minutes	No	Yes
Scholastic Phonics Inventory	Reading-Screener Version	◐	Moderate High	●	◐	—	Individual Group	10 minutes	Computer Scored	No
STAR	Early Literacy	◐	Broad	●	◐	●	Individual Group	10 minutes	Computer Scored	Yes
	Mathematics	●	Broad	●	●	●	Individual Group	10 minutes	Computer Scored	Yes
	Reading	●	Broad	●	●	●	Individual Group	10 minutes	Computer Scored	Yes
STEEP	Oral Reading Fluency	●	Moderate High	●	◐	—	Individual	1 minute	Yes	Yes
TPRI Early Reading Assessment	Reading	●	Moderate Low	◐	○	◐	Individual	2–6 minutes	Yes	Yes

Legend: ● Convincing evidence ◐ Partially convincing evidence ○ Unconvincing evidence — Data unavailable or inadequate
* Added in the 2014 review ** Information updated during the 2014 review

APPENDIX C

INSTITUTIONAL REVIEW BOARD APPROVAL

UNIVERSITY OF
NORTHERN COLORADO

Institutional Review Board

DATE:	April 28, 2017
TO:	Valerie Sherman, M.A.
FROM:	University of Northern Colorado (UNCO) IRB
PROJECT TITLE:	[1063296-2] AN INVESTIGATION OF THE MULTI-TIERED SYSTEM OF SUPPORTS: IMPLEMENTATION PERCEPTIONS AND 3RD GRADE READING ACHIEVEMENT
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVAL/VERIFICATION OF EXEMPT STATUS
DECISION DATE:	April 27, 2017
EXPIRATION DATE:	April 27, 2021

Thank you for your submission of Amendment/Modification materials for this project. The University of Northern Colorado (UNCO) IRB approves this project and verifies its status as EXEMPT according to federal IRB regulations.

Hello Valerie,

Thank you very much for the copy of the Consent from the original research. Everything looks great and your application is approved. Please note that the Consent indicates that the data from the survey is also de-identified before using in the analysis.

Good luck with your research.

Sincerely,

Nancy White, PhD, IRB Reviewer

We will retain a copy of this correspondence within our records for a duration of 4 years.

If you have any questions, please contact Sherry May at 970-351-1910 or Sherry.May@unco.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within University of Northern Colorado (UNCO) IRB's records.

APPENDIX D

MULTI-TIERED SYSTEM OF SUPPORTS IMPLEMENTATION PERCEPTION SURVEY ITEM-LEVEL DESCRIPTIVE STATISTICS

Mean, Standard Deviation, Skewness and Kurtosis Values for Items

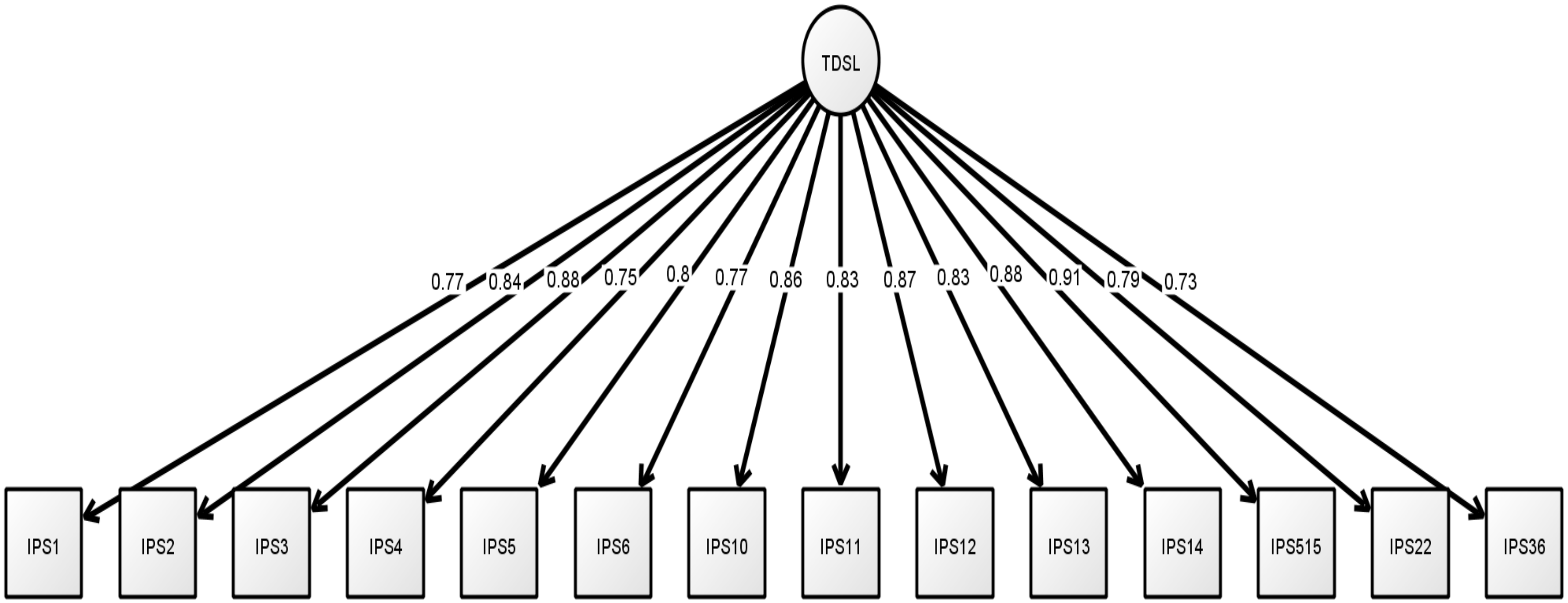
Latent Variable, Item	*n*	Missing	Mean	SD	Skewness	Kurtosis
TDSL						
IPS1	375	144	3.88	0.914	IPS1	375
IPS2	376	143	3.73	0.993	-0.393	-0.594
IPS3	376	143	3.88	1.016	-0.560	-0.413
IPS4	374	145	3.59	1.036	-0.354	-0.491
IPS5	375	144	4.01	0.969	-0.726	-0.214
IPS6	372	147	3.81	1.116	-0.669	-0.412
IPS10	342	177	3.70	1.281	-0.734	-0.569
IPS11	338	181	3.30	1.178	-0.358	-0.720
IPS12	342	177	3.95	1.049	-0.892	0.162
IPS13	341	178	3.82	1.148	-0.772	-0.348
IPS14	342	177	3.54	1.224	-0.501	-0.701
IPS15	336	183	3.67	1.193	-0.628	-0.531
IPS22	321	198	3.97	1.168	-0.430	0.894
IPS36	296	223	4.31	0.759	-0.830	-0.012
DBPS						
IPS7	369	150	3.91	1.020	-0.674	-0.208
IPS8	342	177	3.74	1.104	-0.635	-0.328
IPS9	341	178	3.74	1.078	-0.669	-0.211
IPS16	322	197	3.97	1.003	-0.186	0.937
IPS17	322	197	3.75	0.990	-0.215	0.551
IPS18	322	197	3.83	1.028	-0.560	-0.385
IPS19	321	198	4.11	0.980	-0.902	0.036
IPS20	320	199	4.03	1.122	0.166	1.940
IPS21	322	197	3.78	1.396	0.854	1.981
IPS23	309	210	3.95	0.855	-0.497	-0.064
IPS24	313	206	3.70	0.974	-0.344	-0.495
IPS25	309	210	3.57	1.015	-0.321	-0.576
IPS29	300	219	3.46	1.110	-0.444	-0.499
IPS32	290	229	3.44	1.309	-0.516	-0.809
USPM						
IPS40	291	228	3.84	1.155	-0.838	-0.147
IPS41	281	238	3.44	1.319	-0.420	-0.910
IPS42	286	233	3.87	0.984	-0.760	0.184
IPS43	290	229	3.81	0.951	-0.700	0.169
IPS44	288	231	4.40	0.975	-1.839	3.091
IPS45	289	230	4.43	0.775	-1.373	1.718
IPS46	282	237	4.20	0.954	-1.332	1.702
IPS47	276	243	3.90	1.243	-0.972	-0.043
IPS48	270	249	3.66	1.279	-0.689	-0.588
IPS49	280	239	3.94	1.110	-0.892	0.069
IPS50	272	247	3.18	1.150	-0.196	-0.635

FSCP						
IPS37	297	222	4.24	0.802	-0.822	0.237
IPS38	295	224	4.43	0.809	-1.559	2.468
IPS39	293	226	4.08	0.864	-0.800	0.407
EBP						
IPS26	313	206	3.24	1.096	-0.164	-0.716
IPS27	310	209	3.62	1.093	-0.479	-0.418
IPS28	311	208	3.78	1.053	-0.631	-0.152
LCS						
IPS30	301	218	3.36	1.191	-0.456	-0.654
IPS33	300	219	4.20	0.902	-0.976	0.397
IPS34	297	222	4.11	0.888	-0.774	0.112
IPS1	375	144	3.88	0.914	-0.494	-0.341
TCAP	464	55	71.88	14.30	-.80	.138
%FRL	516	3	.49	.24	.08	-.91
District Size	507	12	1.00	.49	1.77	3.74

Note. *n* = number; SD= Standard Deviation; TDSL =Team Driven Shared Leadership; DBPS = Data-Based Problem Solving; FSCP = Family, School, And Community Partnerships; LCS= Layered Continuum Of Supports; EBP = Evidence-Based Practices; TCAP = school-level 3rd grade reading proficiency-levels from the 2014Transitional Colorado Assessment Program; %FRL= % of students in 2014 at an individual school who qualified for free or reduced lunches. %FRL serves as an indicator of the average socio-economic status of the students within an individual school. District Size = Standardized value for district-level total student enrollment.

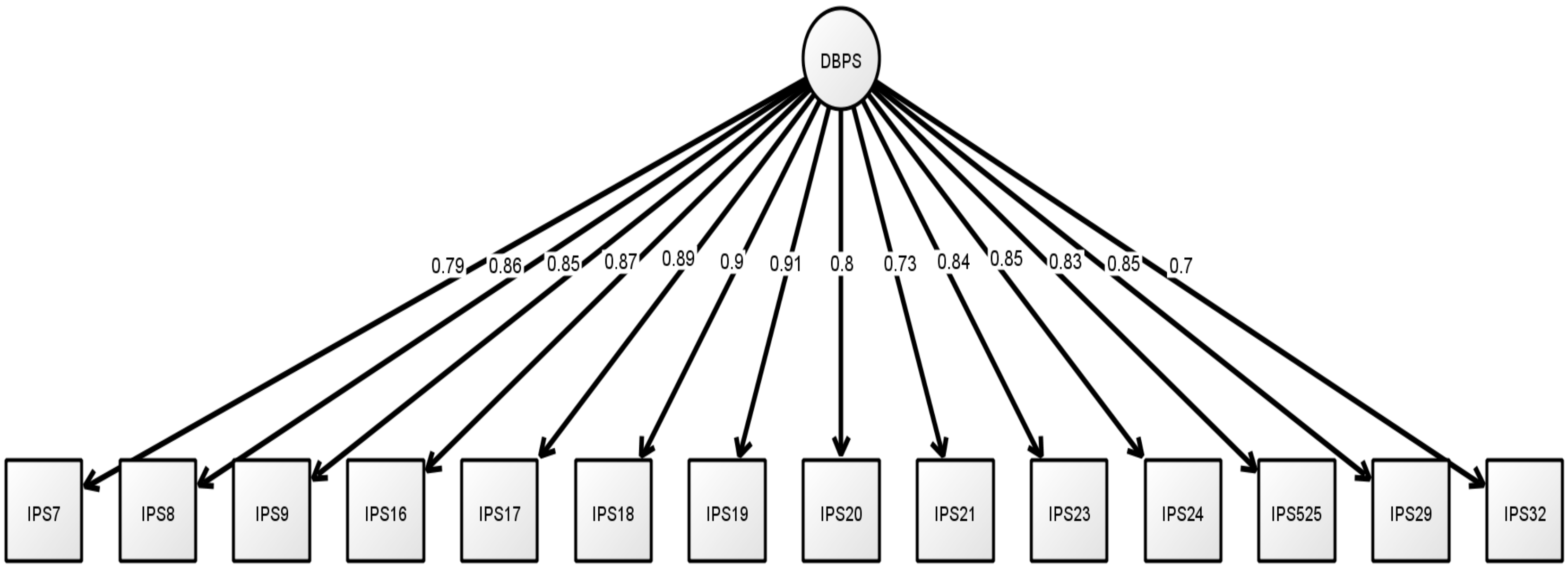
APPENDIX E

CONFIRMATORY FACTOR ANALYSIS: TEAM-DRIVEN SHARED LEADERSHIP


TDSL
0.77
0.84
0.88
0.75
0.8
0.77
0.86
0.83
0.87
0.83
0.88
0.91
0.79
0.73
IPS1
IPS2
IPS3
IPS4
IPS5
IPS6
IPS10
IPS11
IPS12
IPS13
IPS14
IPS515
IPS22
IPS36

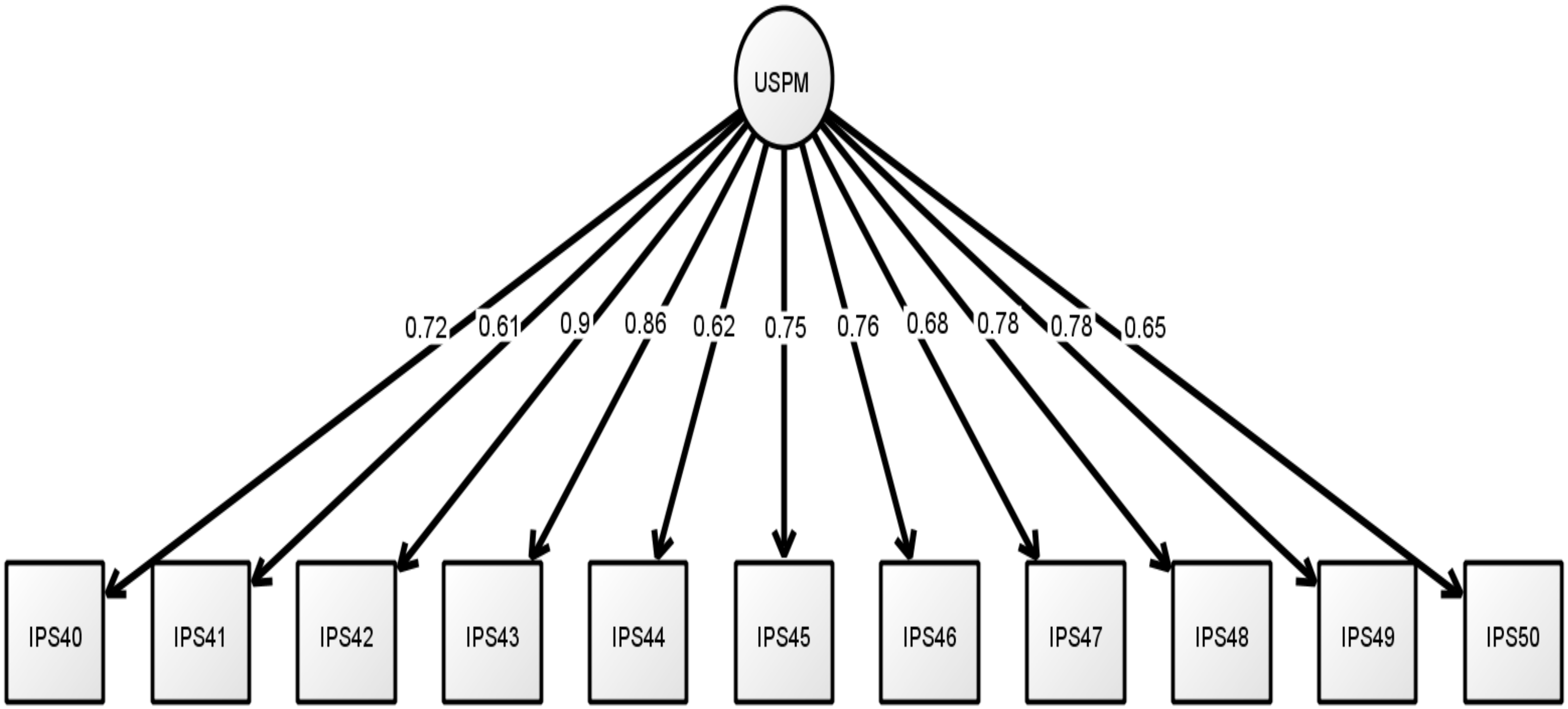
APPENDIX F

CONFIRMATORY FACTOR ANALYSIS: DATA-BASED PROBLEM SOLVING


DBPS
0.79
0.86
0.85
0.87
0.89
0.9
0.91
0.8
0.73
0.84
0.85
0.83
0.85
0.7
IPS7
IPS8
IPS9
IPS16
IPS17
IPS18
IPS19
IPS20
IPS21
IPS23
IPS24
IPS525
IPS29
IPS32

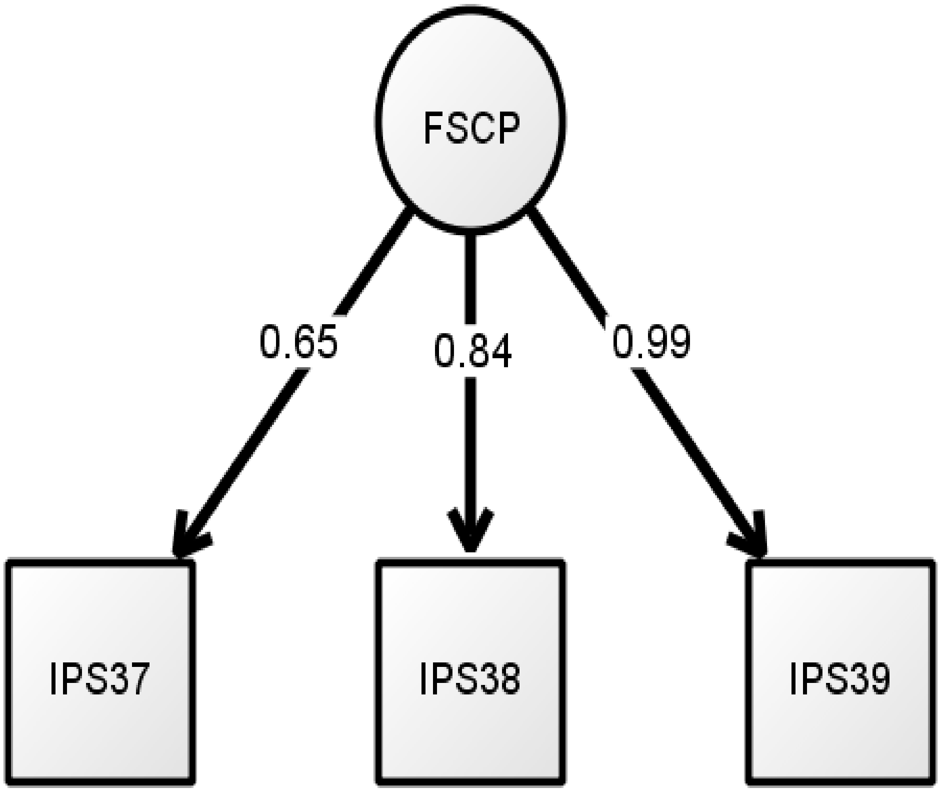
APPENDIX G

CONFIRMATORY FACTOR ANALYSIS: UNIVERSAL SCREENING AND PROGRESS MONITORING


USPM
0.72
0.61
0.9
0.86
0.62
0.75
0.76
0.68
0.78
0.78
0.65
IPS40
IPS41
IPS42
IPS43
IPS44
IPS45
IPS46
IPS47
IPS48
IPS49
IPS50

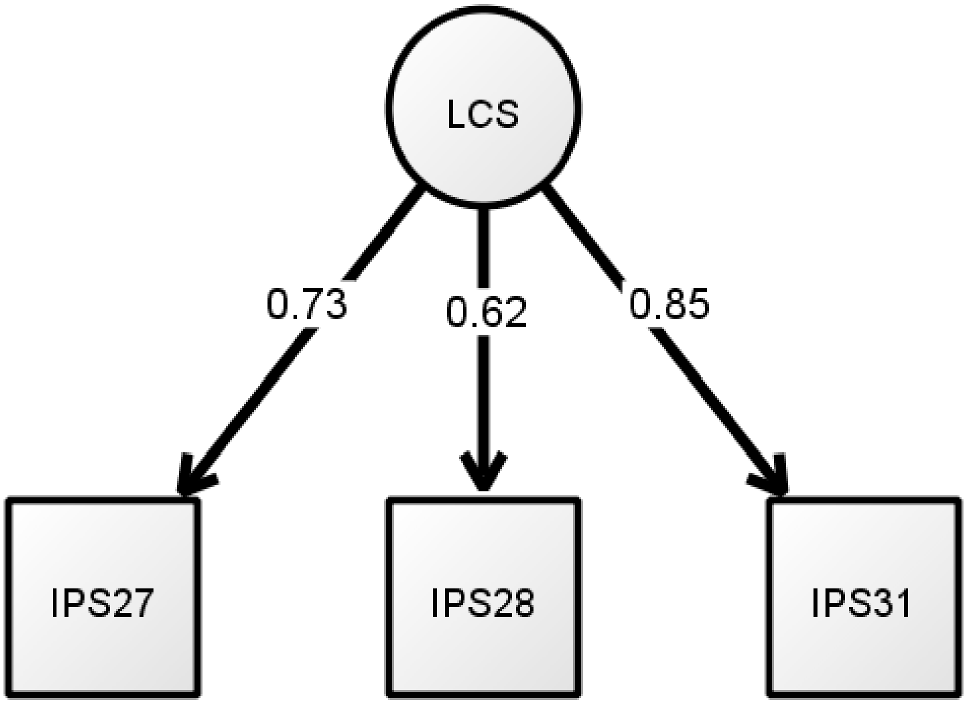
APPENDIX H

CONFIRMATORY FACTOR ANALYSIS: FAMILY, SCHOOL, AND COMMUNITY PARTNERING


FSCP
0.65
0.84
0.99
IPS37
IPS38
IPS39

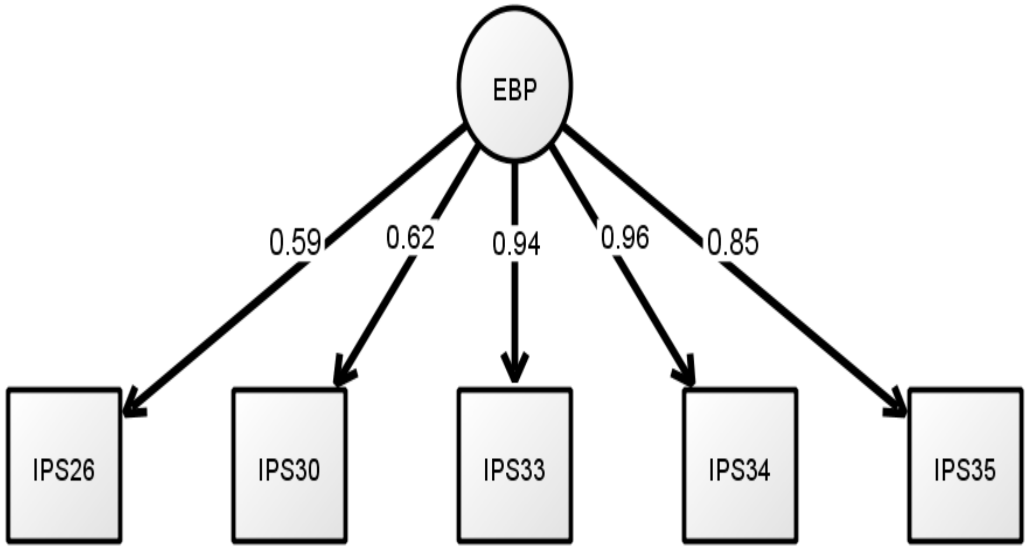
APPENDIX I

CONFIRMATORY FACTOR ANALYSIS: LAYERED CONTINUUM OF SUPPORTS


LCS
0.73
0.62
0.85
IPS27
IPS28
IPS31

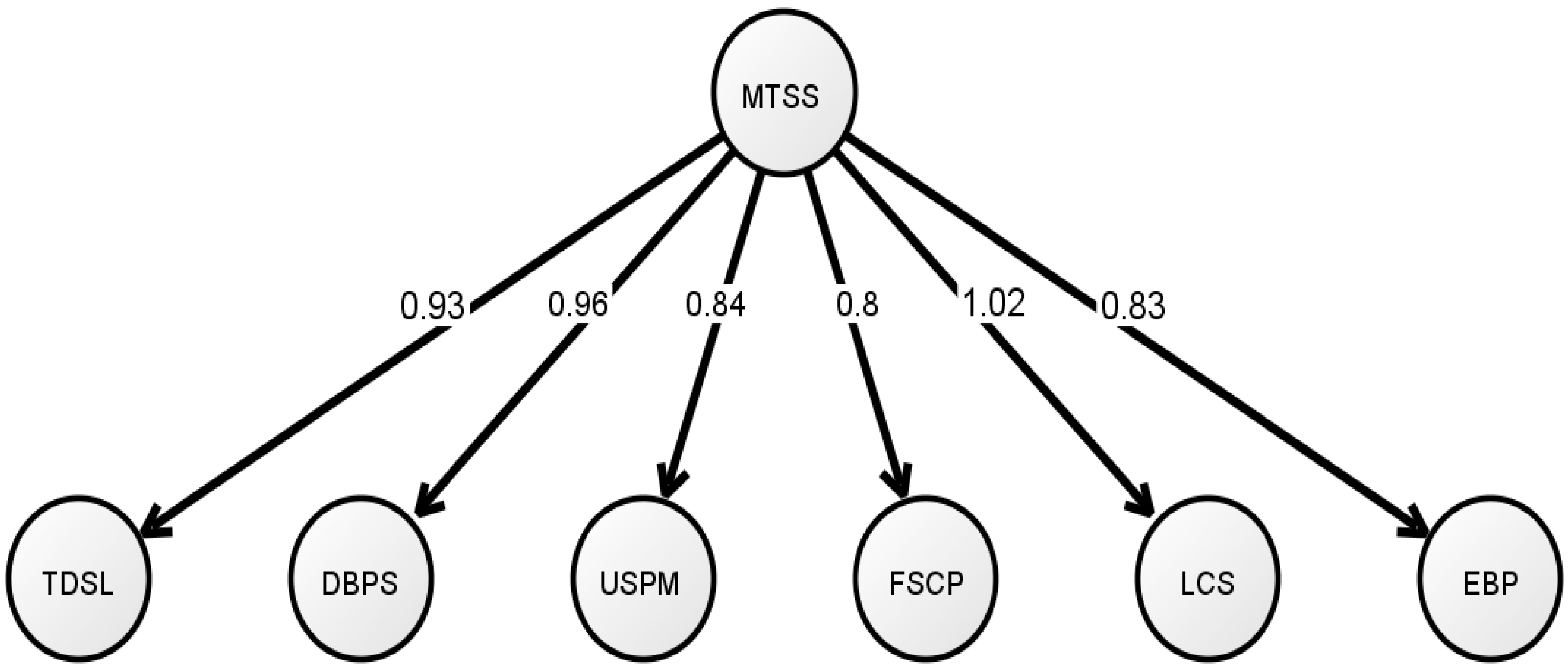
APPENDIX J

CONFIRMATORY FACTOR ANALYSIS: EVIDENCE-BASED PRACTICES


EBP
0.59
0.62
0.94
0.96
0.85
IPS26
IPS30
IPS33
IPS34
IPS35

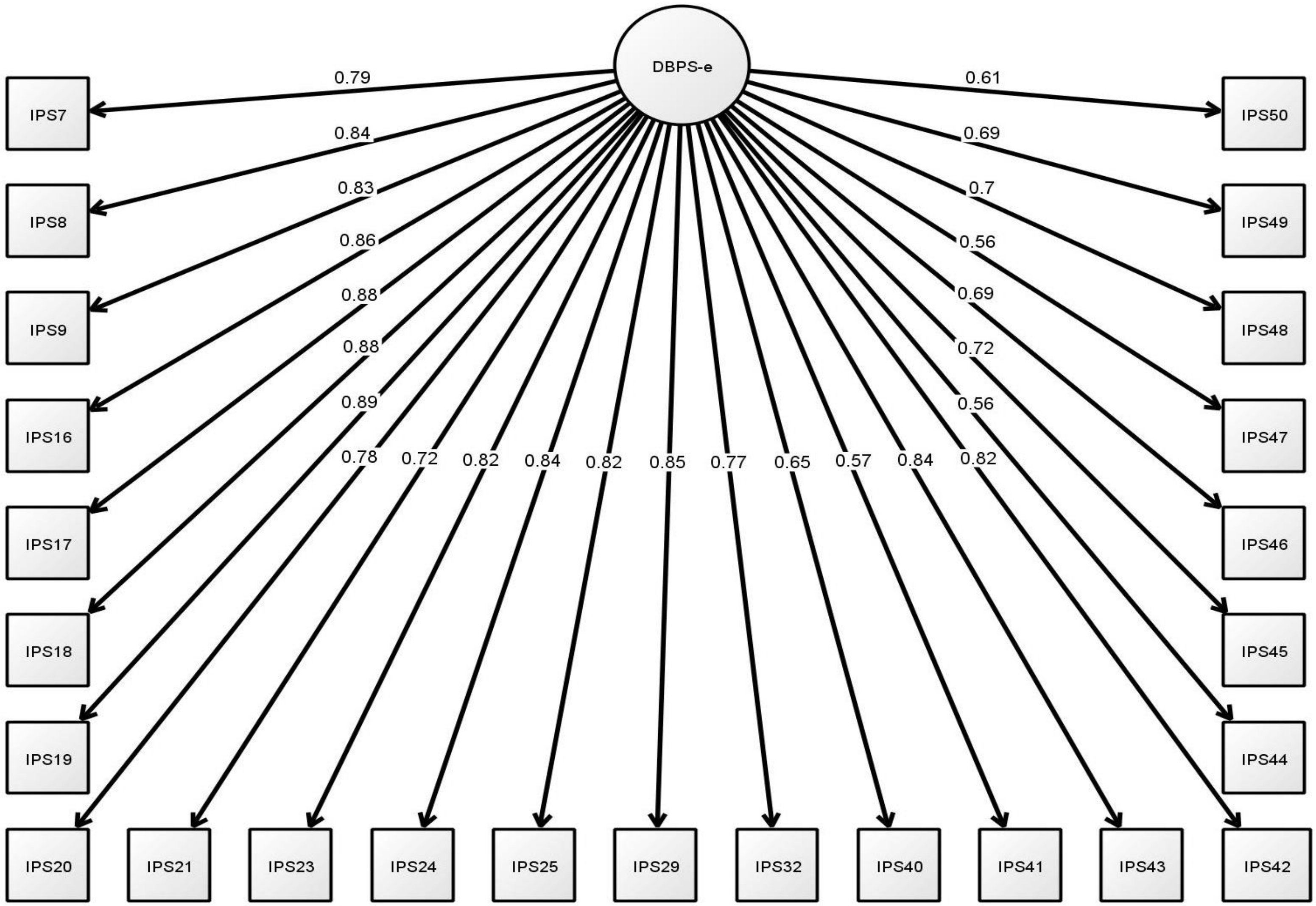
APPENDIX K

SIX-FACTOR ENDOGENOUS MULTI-TIERED SYSTEM OF SUPPORTS MODEL


MTSS
0.93
0.96
0.84
0.8
1.02
0.83
TDSL
DBPS
USPM
FSCP
LCS
EBP

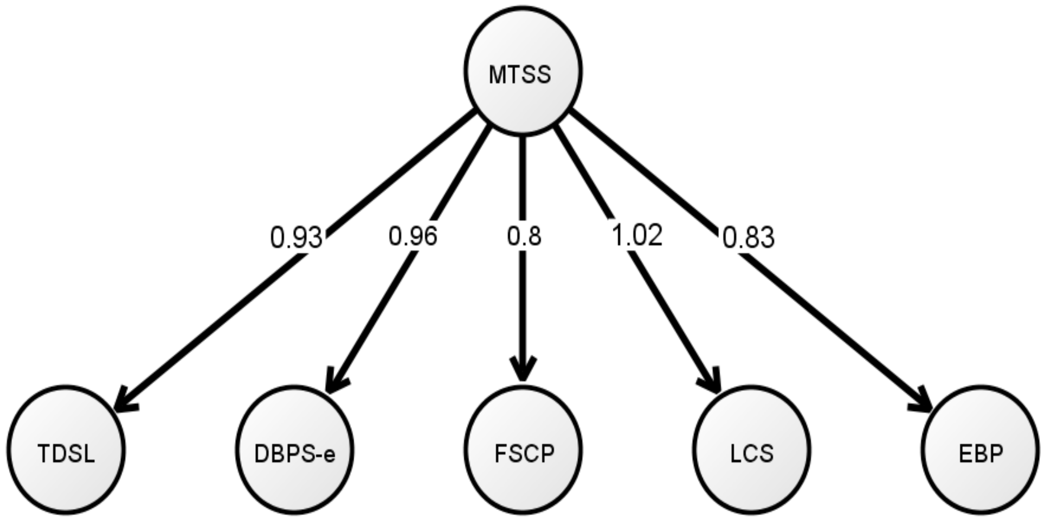
APPENDIX L

CONFIRMATORY FACTOR ANALYSIS: EXPANDED DATA-BASED PROBLEM SOLVING


DBPS-e
IPS7
IPS8
IPS9
IPS16
IPS17
IPS18
IPS19
IPS20
IPS21
IPS23
IPS24
IPS25
IPS29
IPS32
IPS40
IPS41
IPS43
IPS42
IPS44
IPS45
IPS46
IPS47
IPS48
IPS49
IPS50
0.79
0.84
0.83
0.86
0.88
0.88
0.89
0.78
0.72
0.82
0.84
0.82
0.85
0.77
0.65
0.57
0.84
0.82
0.56
0.72
0.69
0.56
0.7
0.69
0.61

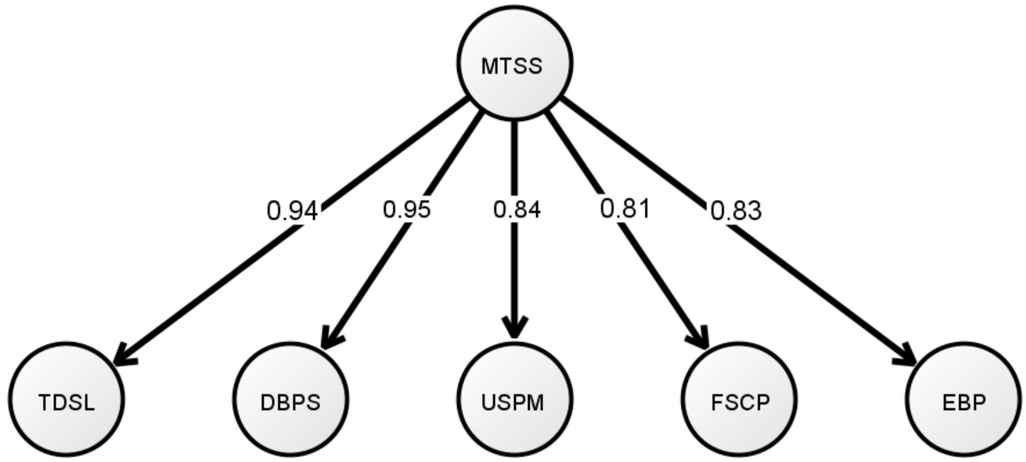
APPENDIX M

FIVE-FACTOR ENDOGENOUS MULTI-TIERED SYSTEM OF SUPPORTS MODEL


MTSS
0.93
0.96
0.8
1.02
0.83
TDSL
DBPS-e
FSCP
LCS
EBP

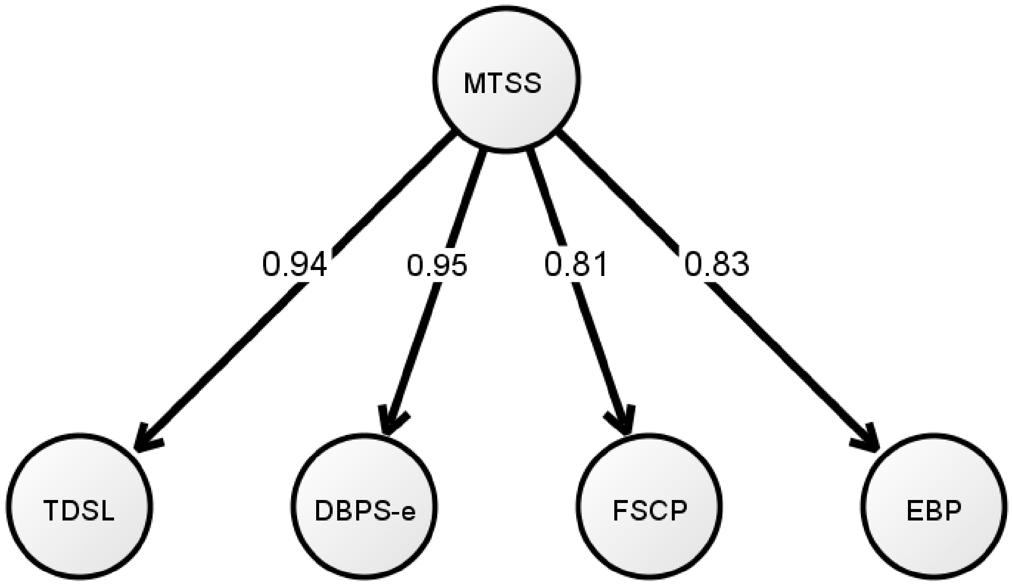
APPENDIX N

REVISED FIVE-FACTOR ENDOGENOUS MULTI-TIERED SYSTEM OF SUPPORTS MODEL


MTSS
0.94
0.95
0.84
0.81
0.83
TDSL
DBPS
USPM
FSCP
EBP

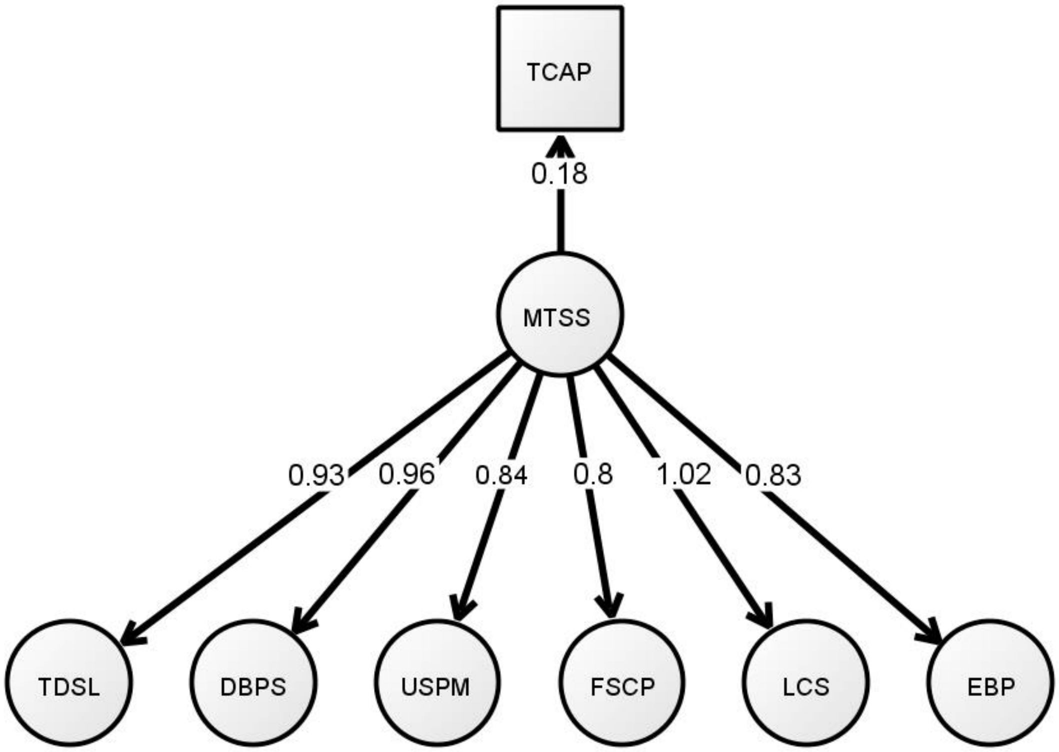
APPENDIX O

FOUR-FACTOR ENDOGENOUS MULTI-TIERED SYSTEM OF SUPPORTS MODEL


MTSS
0.94
0.95
0.81
0.83
TDSL
DBPS-e
FSCP
EBP

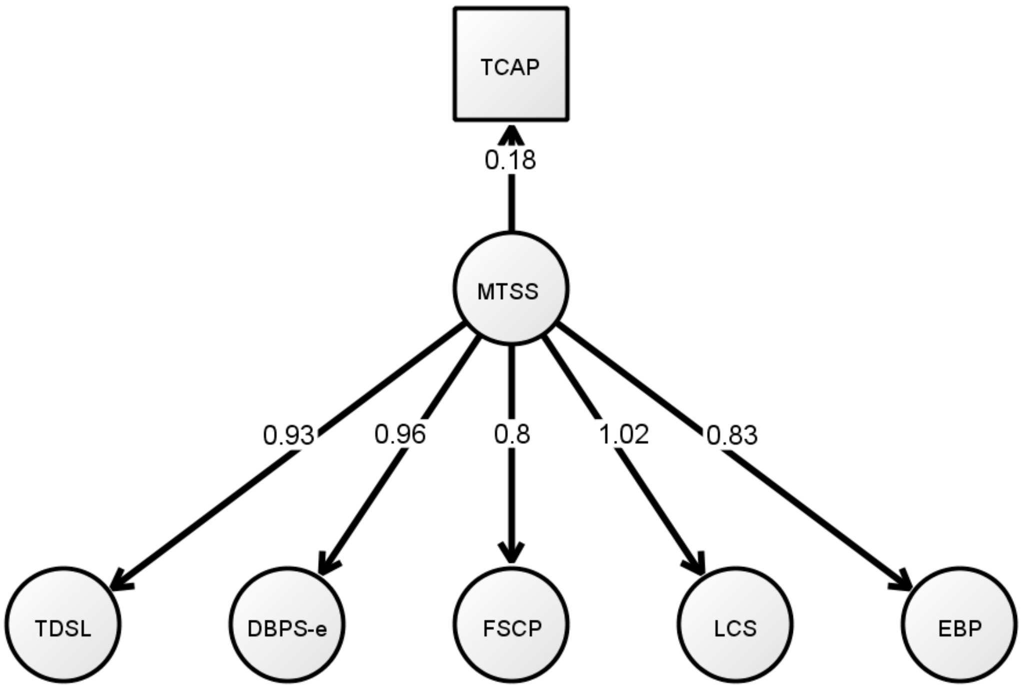
APPENDIX P

SIX-FACTOR HIGHER-ORDER STRUCTURAL MODEL


TCAP
0.18
MTSS
0.93
0.96
0.84
0.8
1.02
0.83
TDSL
DBPS
USPM
FSCP
LCS
EBP

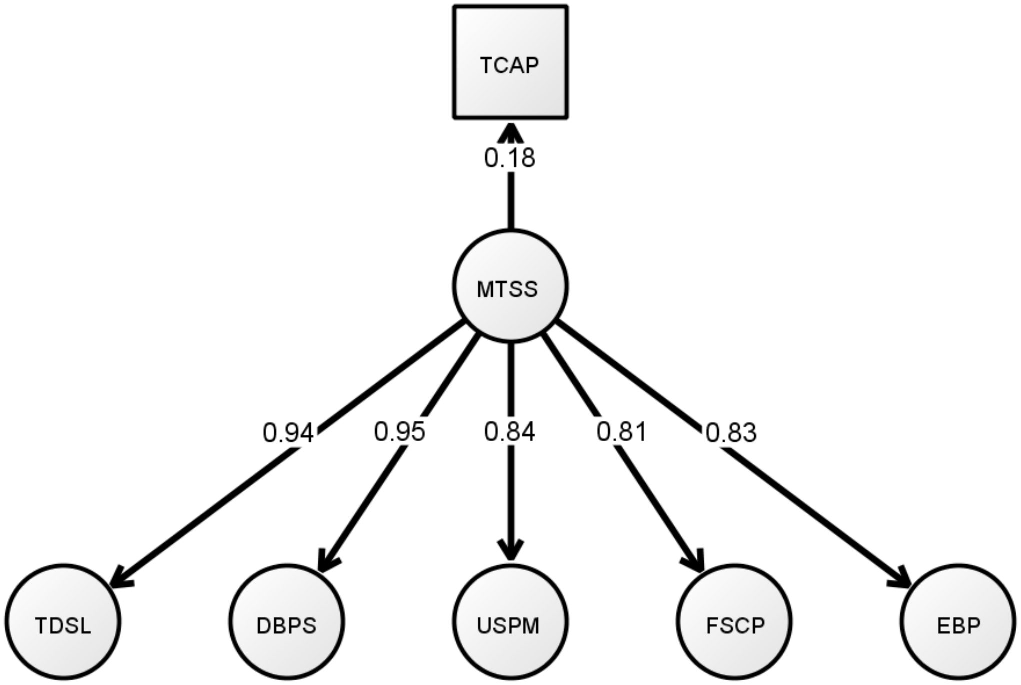
APPENDIX Q

FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL


TCAP
0.18
MTSS
0.93
0.96
0.8
1.02
0.83
TDSL
DBPS-e
FSCP
LCS
EBP

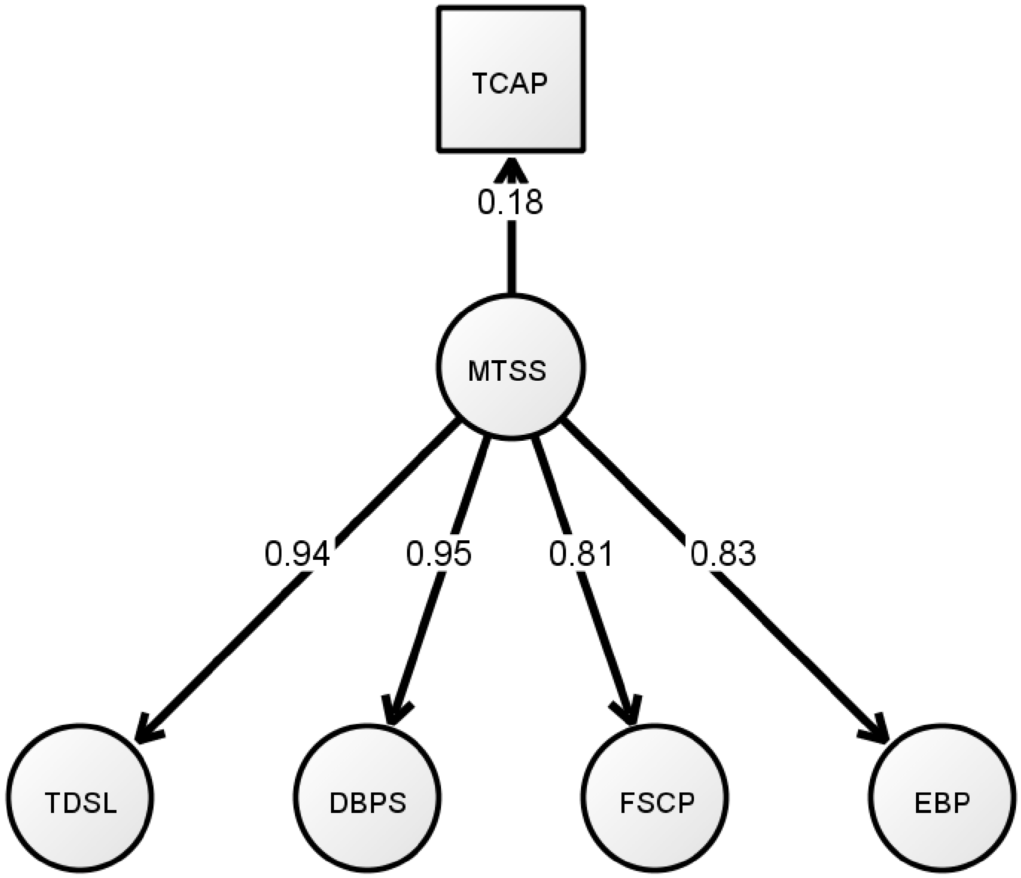
APPENDIX R

REVISED FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL


TCAP
0.18
MTSS
0.94
0.95
0.84
0.81
0.83
TDSL
DBPS
USPM
FSCP
EBP

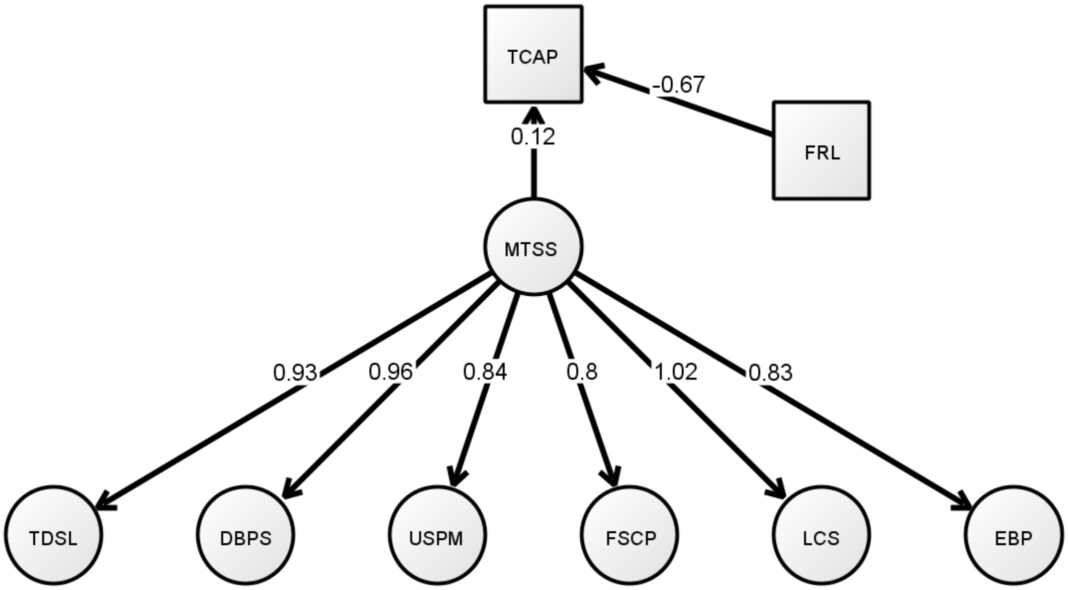
APPENDIX S

FOUR-FACTOR HIGHER-ORDER STRUCTURAL MODEL


TCAP
0.18
MTSS
0.94
0.95
0.81
0.83
TDSL
DBPS
FSCP
EBP

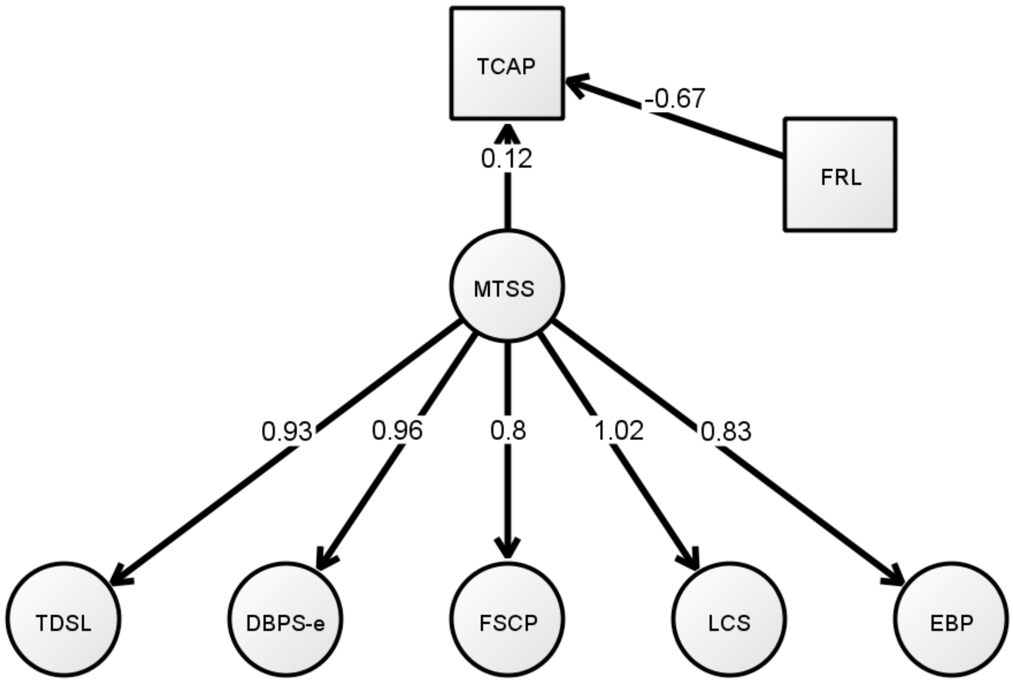
APPENDIX T

SIX-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH


TCAP
-0.67
FRL
0.12
MTSS
0.93
0.96
0.84
0.8
1.02
0.83
TDSL
DBPS
USPM
FSCP
LCS
EBP

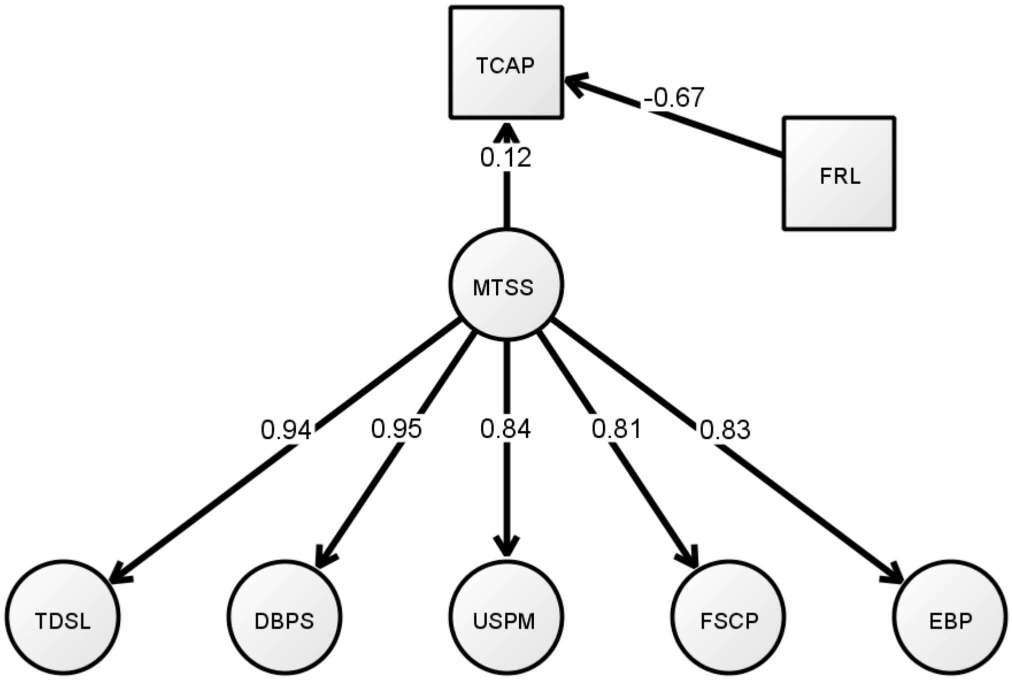
APPENDIX U

FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODELWITH FREE AND REDUCED LUNCH


TCAP
-0.67
FRL
0.12
MTSS
0.93
0.96
0.8
1.02
0.83
TDSL
DBPS-e
FSCP
LCS
EBP

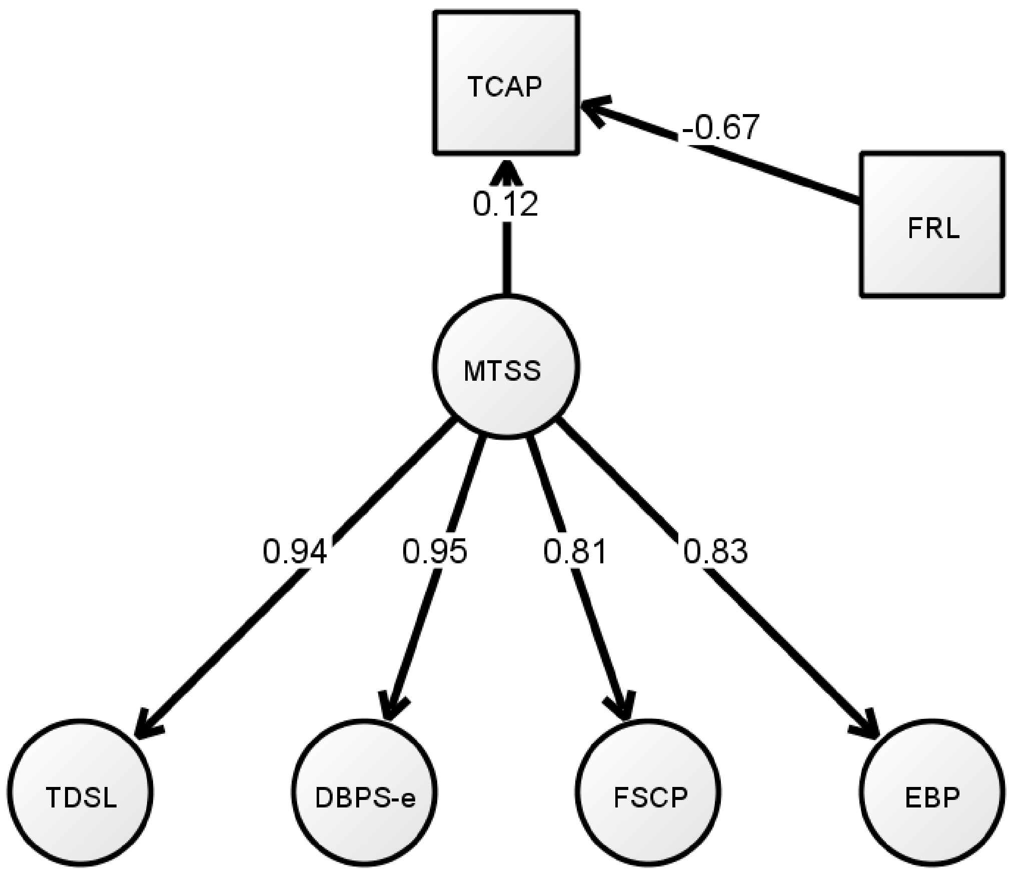
APPENDIX V

REVISED FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH


TCAP
-0.67
FRL
0.12
MTSS
0.94
0.95
0.84
0.81
0.83
TDSL
DBPS
USPM
FSCP
EBP

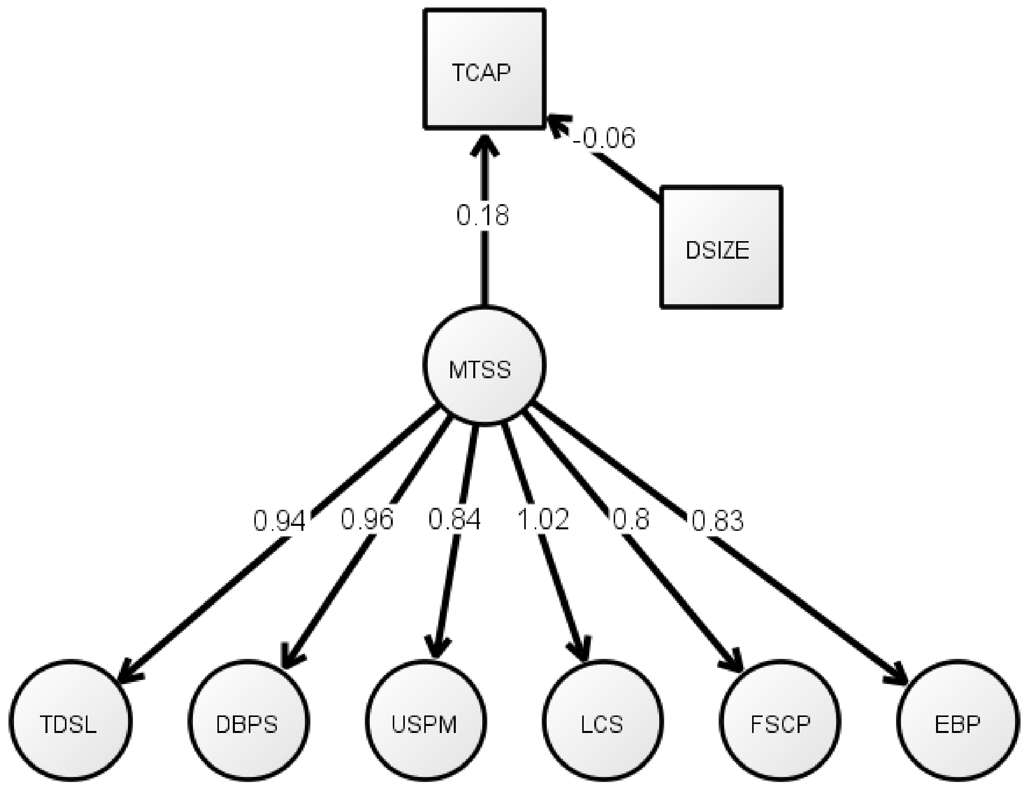
APPENDIX W

FOUR-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH


TCAP
-0.67
FRL
0.12
MTSS
0.94
0.95
0.81
0.83
TDSL
DBPS-e
FSCP
EBP

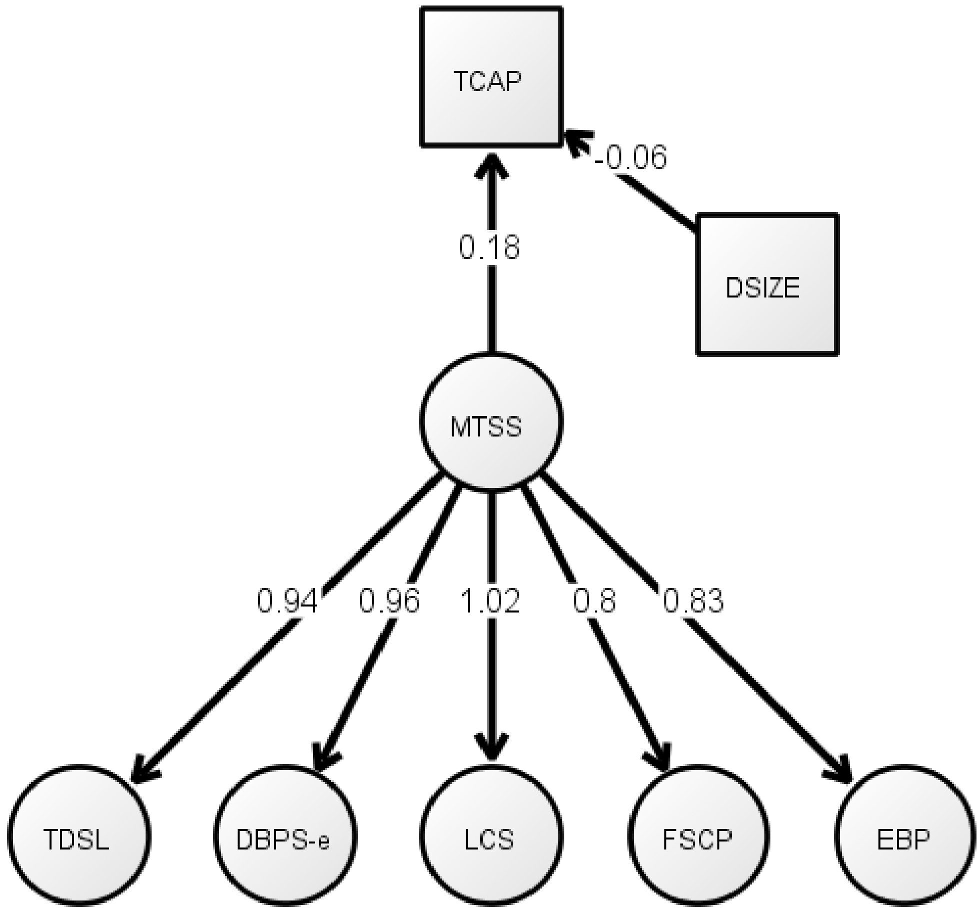
APPENDIX X

SIX-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH DISTRICT SIZE


TCAP
-0.06
0.18
DSIZE
MTSS
0.94
0.96
0.84
1.02
0.8
0.83
TDSL
DBPS
USPM
LCS
FSCP
EBP

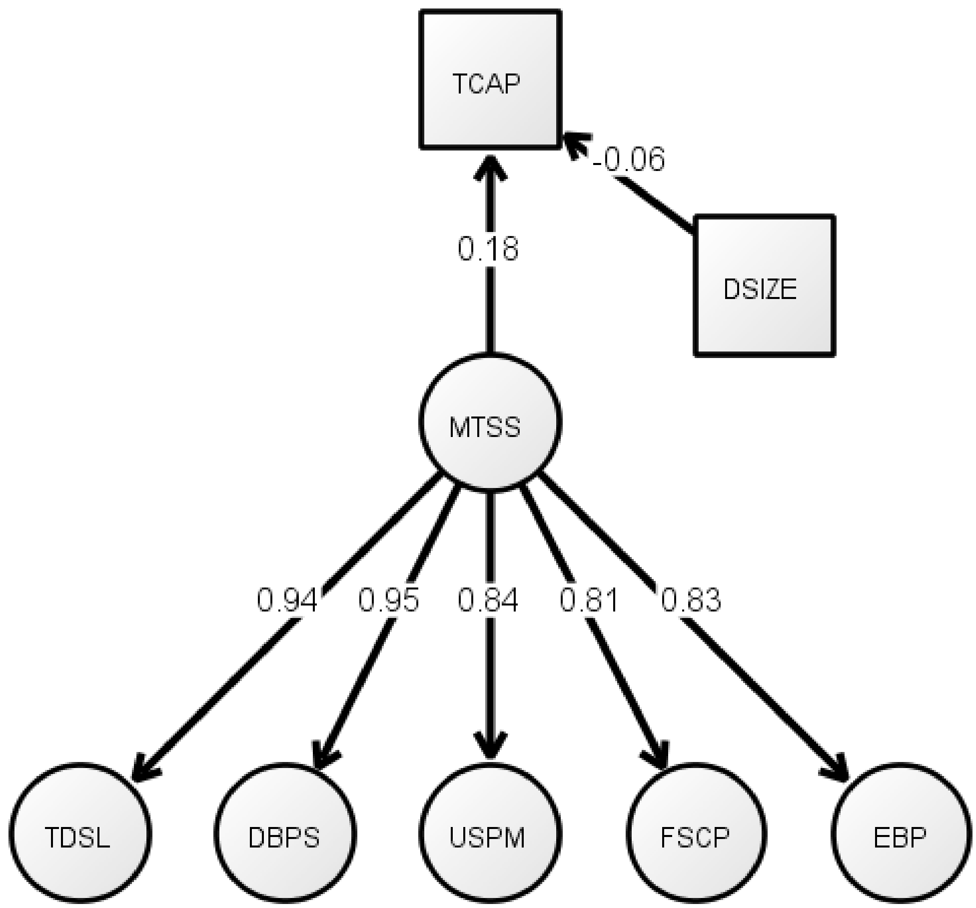
APPENDIX Y

FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH DISTRICT SIZE


TCAP
-0.06
0.18
DSIZE
MTSS
0.94
0.96
1.02
0.8
0.83
TDSL
DBPS-e
LCS
FSCP
EBP

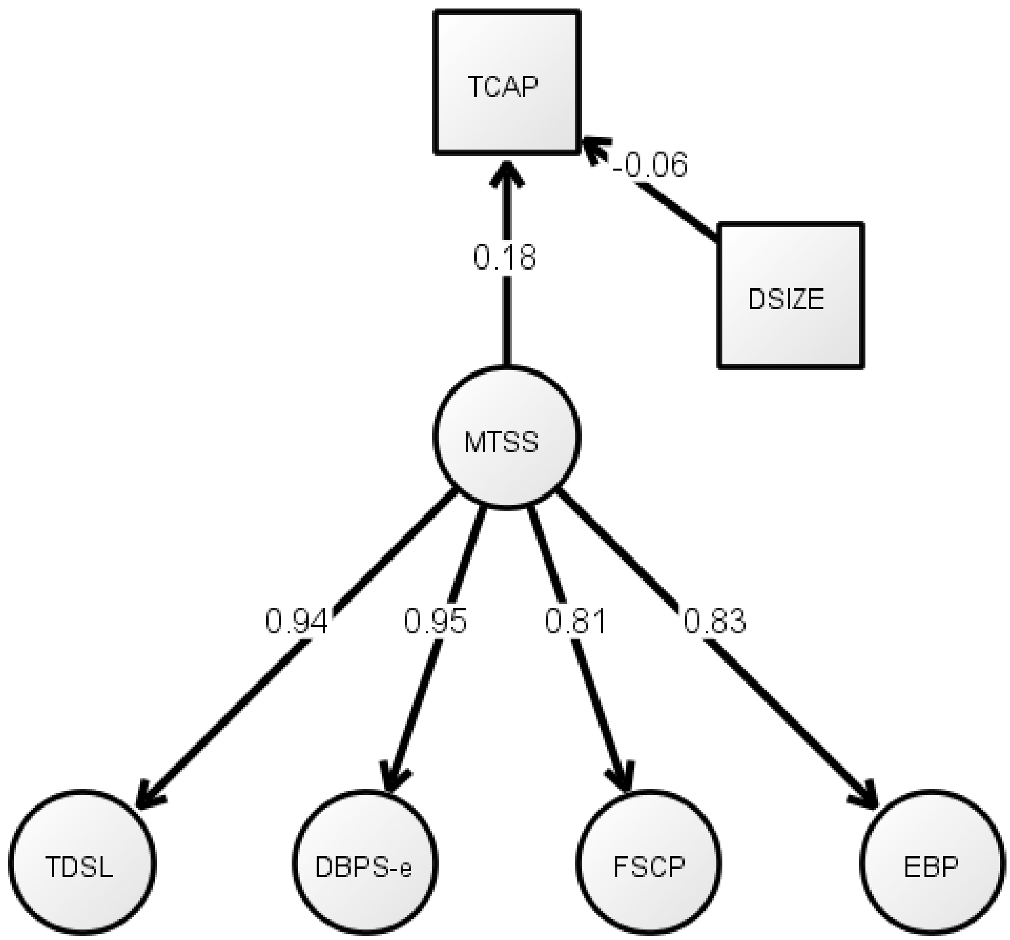
APPENDIX Z

REVISED FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH DISTRICT SIZE


TCAP
-0.06
0.18
DSIZE
MTSS
0.94
0.95
0.84
0.81
0.83
TDSL
DBPS
USPM
FSCP
EBP

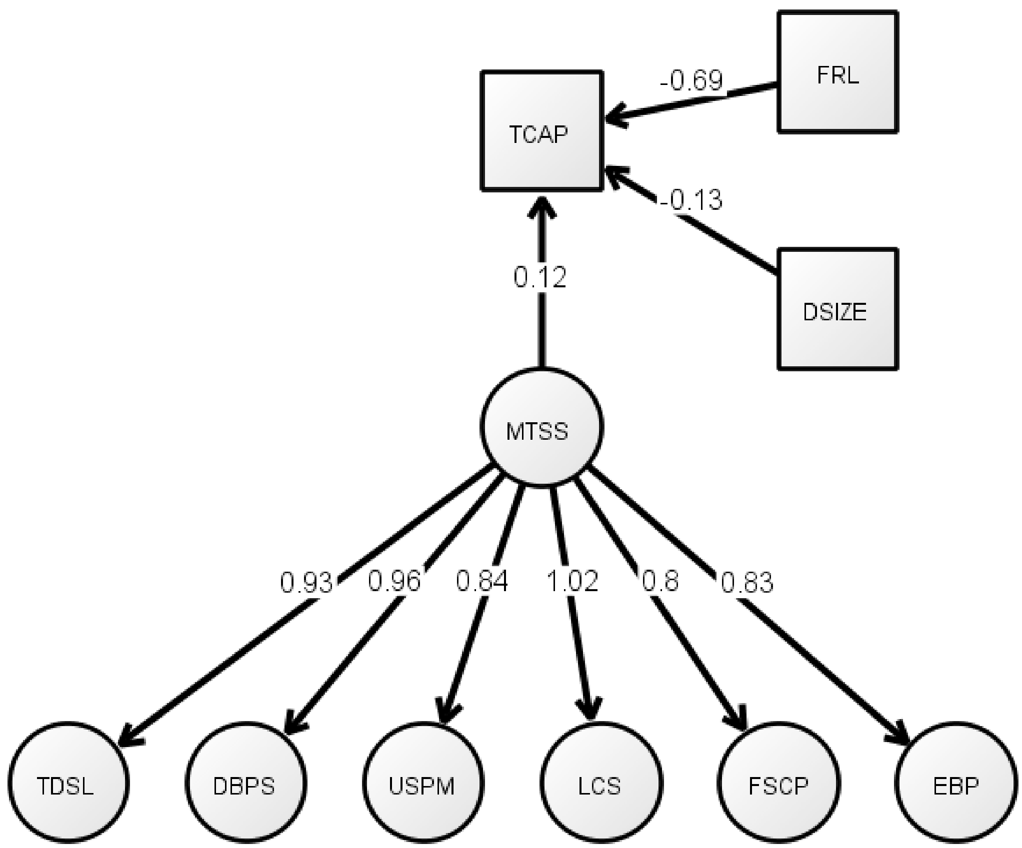
APPENDIX AA

FOUR-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH DISTRICT SIZE


TCAP
-0.06
0.18
DSIZE
MTSS
0.94
0.95
0.81
0.83
TDSL
DBPS-e
FSCP
EBP

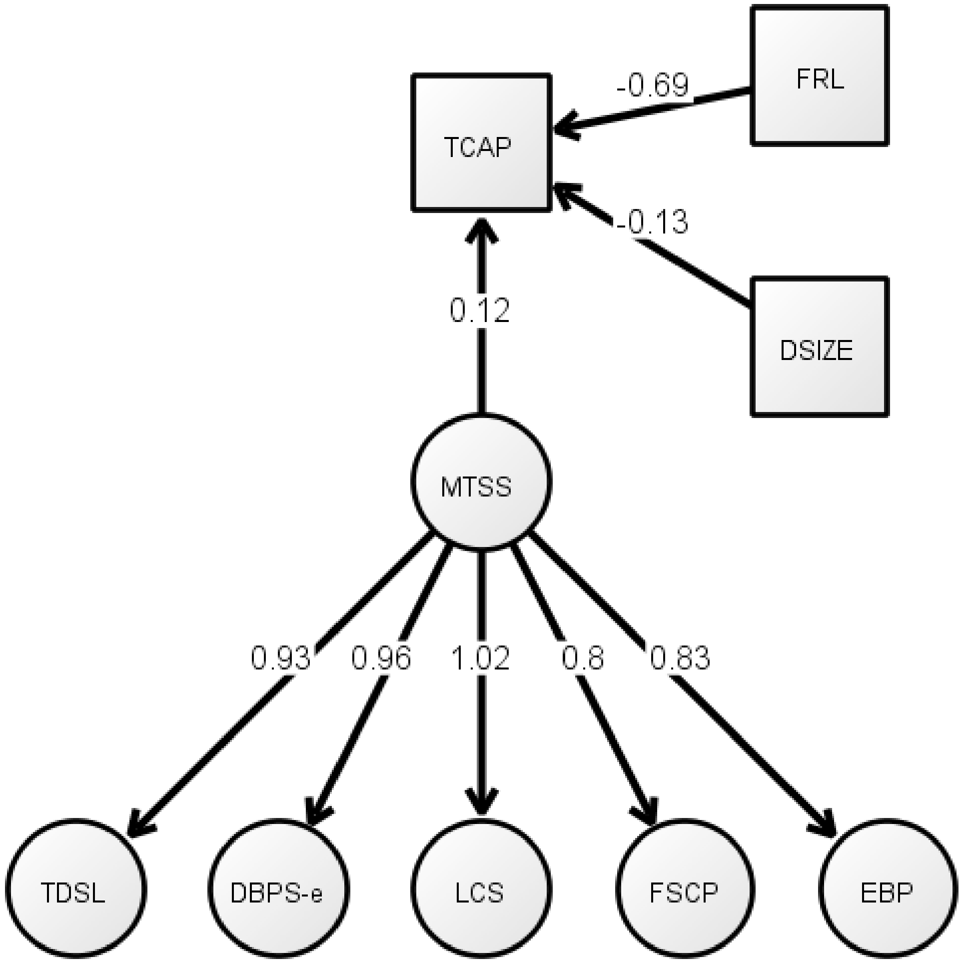
APPENDIX AB

SIX-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH AND DISTRICT SIZE


TCAP
FRL
-0.69
-0.13
DSIZE
0.12
MTSS
0.93
0.96
0.84
1.02
0.8
0.83
TDSL
DBPS
USPM
LCS
FSCP
EBP

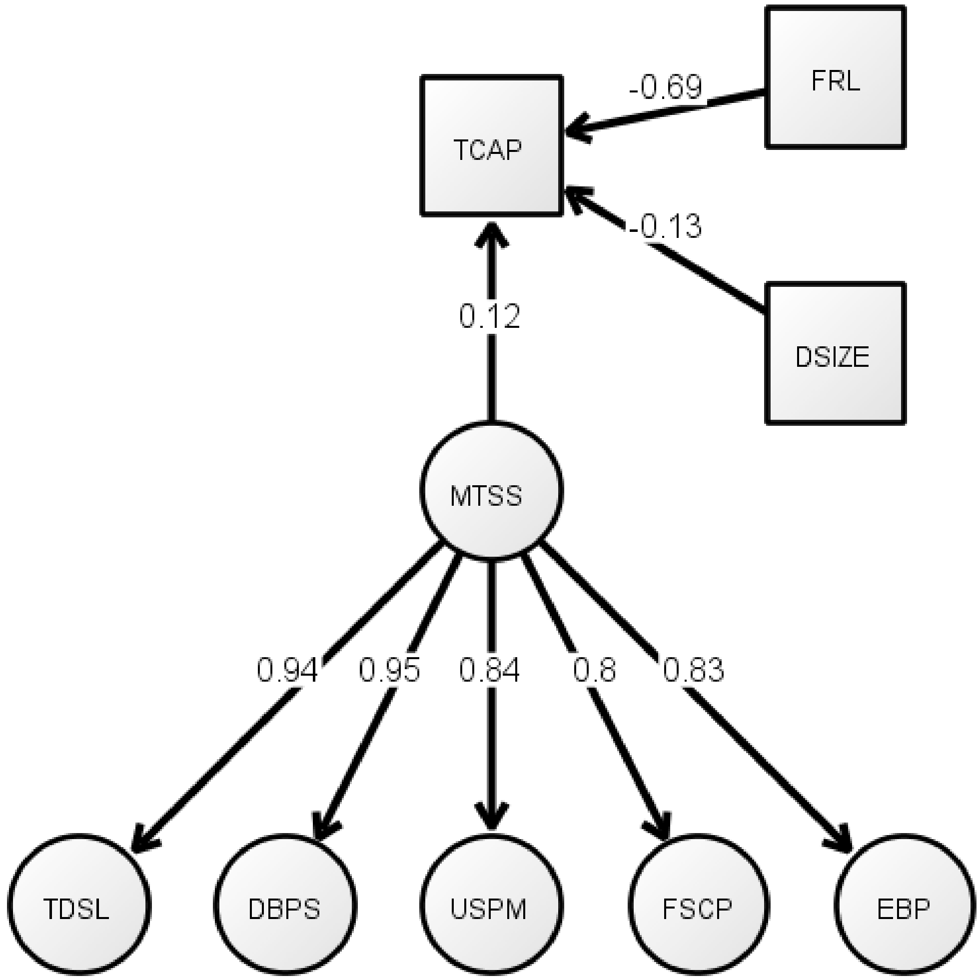
APPENDIX AC

FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH AND DISTRICT SIZE


TCAP
FRL
-0.69
-0.13
DSIZE
0.12
MTSS
0.93
0.96
1.02
0.8
0.83
TDSL
DBPS-e
LCS
FSCP
EBP

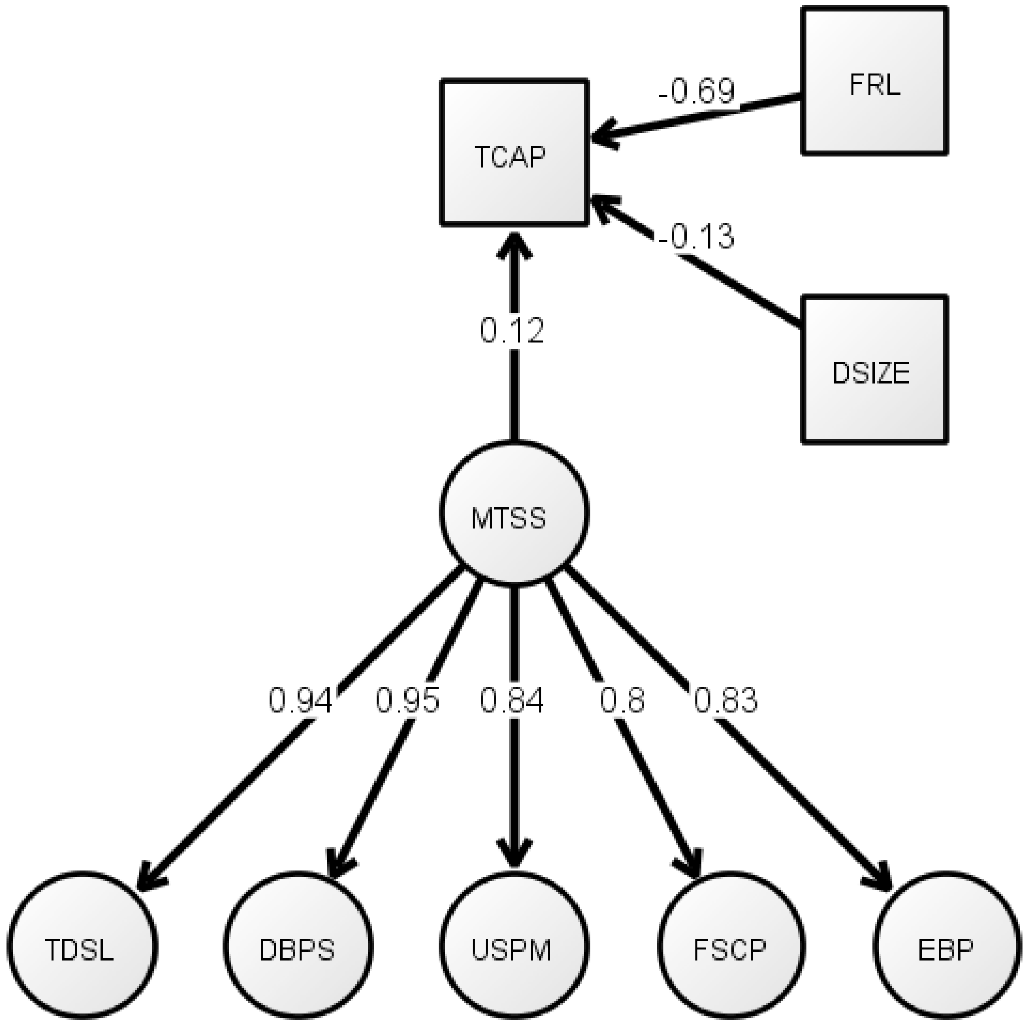
APPENDIX AD

REVISED FIVE-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH AND DISTRICT SIZE


TCAP
FRL
-0.69
-0.13
DSIZE
0.12
MTSS
0.94
0.95
0.84
0.8
0.83
TDSL
DBPS
USPM
FSCP
EBP

APPENDIX AE

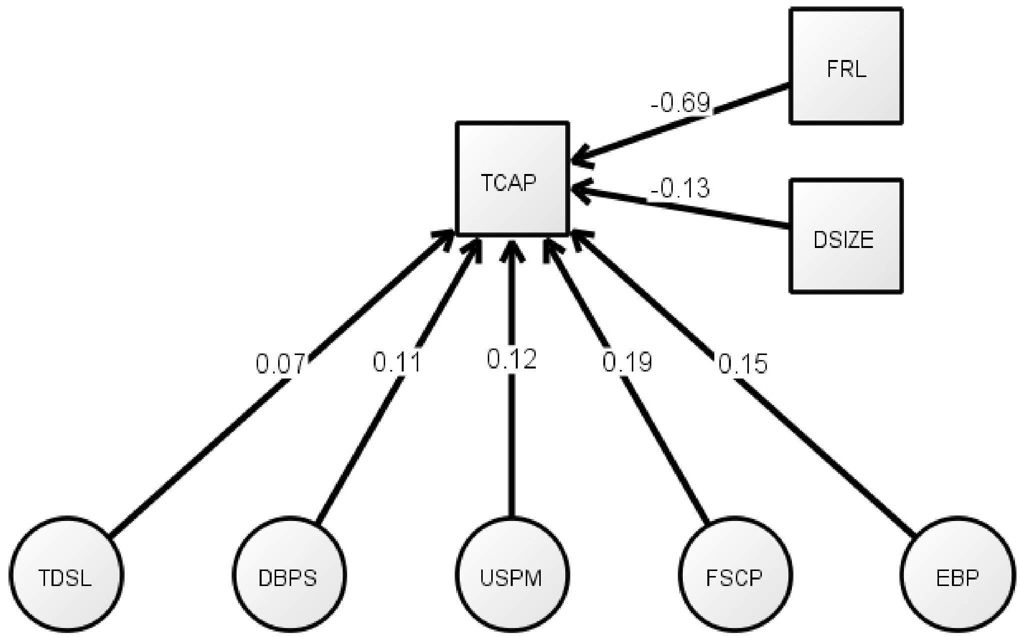
FOUR-FACTOR HIGHER-ORDER STRUCTURAL MODEL WITH FREE AND REDUCED LUNCH AND DISTRICT SIZE



FRL
-0.69
TCAP
-0.13
DSIZE
0.12
MTSS
0.94
0.95
0.84
0.8
0.83
TDSL
DBPS
USPM
FSCP
EBP

APPENDIX AF

DIRECT EFFECTS OF MULTI-TIERED SYSTEM OF SUPPORTS LATENT FACTORS ON TRANSITIONAL COLORADO ASSESSMENT PROGRAM


FRL
-0.69
TCAP
-0.13
DSIZE
0.07
0.11
0.12
0.19
0.15
TDSL
DBPS
USPM
FSCP
EBP