Learning Disabilities: A Contemporary Journal 10(2), 33-52, 2012

The School Implementation Scale: Measuring Implementation in Response to Intervention Models

Amy Gaumer Erickson
Pattie M. Noonan
University of Kansas

Ronda Jenson
University of Missouri, Kansas City

Models of response to intervention (RTI) have been widely developed and implemented and have expanded to include integrated academic/behavior RTI models. Until recently, evaluation of model effectiveness has focused primarily on student-level data, but additional measures of treatment integrity within these multi-tiered models are emerging to facilitate understanding of implementation of essential elements for effective school systems that produce student-level academic and behavioral success. The valid and reliable School Implementation Scale*, a 33-item cost-effective online implementation measure, was designed for all school staff to complete and utilize in school improvement efforts. Moderate correlations were found between scale results and reading and writing achievement of students with disabilities.*

Across the nation, schools are implementing tiered levels of academic and behavioral support (Kalberg, Lane, & Menzies, 2010; Spaulding, Horner, May, & Vincent, 2008). The implementation components of such models are typically well-defined, but treatment integrity across grade levels and tiers has proven difficult to measure (Sanetti & Kratochwill, 2009). Generally, response to intervention (RTI) measures include student-level data (e.g., test scores, discipline referrals). To measure treatment integrity, many multi-tiered interventions include fidelity measures completed by school leadership teams. These measures identify perceptions of a small group of educators, but often fail to address school-wide implementation among all instructional staff. Furthermore, the most expansive evaluations of tiered models incorporate onsite observations, interviews, and focus groups. These onsite measures produce an abundance of data but are often cost-prohibitive and time-consuming (Cohen, Kincaid, & Childs, 2007). While current RTI evaluation procedures elicit valuable information, they have limitations in their scope; specifically, they do not measure the academic and behavioral intervention implementation of all school instructional staff (i.e., school-wide individuals who actually implement the tiered components with students).

To address these limitations, the *School Implementation Scale* was developed to encompass the evidence-based essential elements of effective school systems. In one state's integrated model of academic and behavioral support pilot project, all school instructional staff in 22 schools (14 elementary, four middle, and four high schools) completed the survey annually over two consecutive years, and summary re-

ports were provided to the school principals. School leadership teams identified this data as invaluable to informing action planning goals and activities. Furthermore, this survey has been found to be highly reliable, and results produced moderate correlations with reading and writing proficiency of students with disabilities.

Literature Review

Overview of Tiered Models

While many tiered models of academic and behavioral support are currently being implemented in schools across the United States, the definition of RTI provided by Mellard, Stern and Woods (2011) provides a comprehensive schema: "Response to intervention (RTI) is widely used as a framework for providing high quality instruction and interventions that are matched to students' needs" (p. 1). Well-defined RTI models provide multi-tiered supports to prevent academic and behavioral difficulties as well as to address existing academic and behavioral difficulties. Multiple models incorporating universal screening, progress monitoring, and tiered interventions have been coined "Response to Intervention," "positive behavior supports," or "integrated models." Originally, RTI focused on reading achievement and the identification of students with learning disabilities, while positive behavioral support models focused on behavioral expectations; since then, research has acknowledged the interrelatedness of academic and behavioral performance (Algozzine, Wang & Violette, 2011; Kalberg et al., 2010; Lane, Kalberg, & Menzies, 2009).

Schools have found it to be difficult to implement multiple separate reading-focused and behavior-focused models simultaneously; therefore, many states are developing integrated academic/behavior RTI models. These integrated multi-tiered models include the essential elements of both academic and behavioral interventions while continuing to follow the overarching definition of RTI (Mellard et al., 2011). For example, the Kansas Multi-Tier System of Supports (MTSS) is identified as a "coherent continuum of evidence-based, system-wide practices to support a rapid response to academic and behavioral needs, with frequent data-based monitoring for instructional decision making to empower each Kansas student to achieve to high standards" (Kansas Department of Education, 2011, p. 3). In Florida, the Multi-tiered System of Supports (MTSS) "represents the integration of RTI for academics and RTI for behavior into a unified model of service delivery that recognizes the reciprocal influence academic performance and social/emotional/behavioral performance have on each other" (Florida's Positive Behavior Support Project, 2011, p. 2). Michigan's Integrated Behavior and Learning Support Initiative is described as a systems approach to "develop support systems and sustained implementation of data-driven problem solving models in schools to help students become better readers with social skills necessary for success" (Goodman, 2006, p. 1). The Missouri Integrated Model is similar in that it promotes a process for improving student achievement and transforming school culture with the "goal of creating collaborative and effective schools where parents, community members, and school staff work together in making data-driven decisions to ensure positive social and educational benefit for all students" (Missouri Department of Education, 2010, p. 2). All of these integrated academic/behavior RTI models are designed as proactive approaches to multi-tiered student support.

Implementation Measures within Tiered Models

The success of school-wide multi-tiered models has been found to be dependent upon educators' knowledge and implementation of effective interventions (Barton-Arwood, Morrow, Lane, & Jolivette, 2005; Noell & Gansle, 2006; Sugai & Horner, 2002). Therefore, evaluation of integrated academic/behavior RTI models should include measures of fidelity of implementation: delivery of instruction and support as they were designed to be delivered (Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Burns, Appleton, & Stehouwer, 2005; Gresham, MacMillan, Boebe-Frankenberger, & Bocian, 2000). This concept has also been referred to as treatment integrity, or "the extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention" (Sanetti & Kratochwill, 2009, p. 448). Fidelity/integrity measures typically address four dimensions: (1) content, (2) process, (3) quality, and (4) quantity (Sanetti & Kratochwill, 2009). Within the educational context of indirect and direct service delivery, two levels of treatment integrity have been further conceptualized by Noell (2008): consultation procedural integrity (i.e., training and consultation by a coach or expert) and treatment plan implementation (i.e., school staff delivery of the intervention). Within this section, we focus specifically on measures related to treatment plan implementation.

Content measures have been typically identified as the procedural documentation and artifacts commonly used by school leadership teams to guide implementation. Examples include: needs assessment measures, process checklists, action planning templates to focus on identified areas of school improvement, and progress monitoring of school improvement action plans. Together these documents outline the core components of the RTI model as well as a plan for delivery of the model. Within RTI, these components include tiered structures, universal screening, and progress monitoring (Mellard et. al, 2011).

Process evaluation often has included process checklists measuring the percentage of the model core features that are in place. These measures have typically contained yes/no response options or rating scales for the basic components of the RTI model. For example, an item rated on the *Team Implementation Checklist* states "Team has regular meeting schedule and effective operating procedures," and can be rated as "not started," "in progress," or "in place" (Sugai, Horner, Lewis-Palmer, 2001).

Quality and quantity measures consider the extent to which the model is implemented as intended across the school environment. These fidelity of implementation measures often include interviews, observations, and self-assessments. Numerous measures have been developed to support implementation of school-wide positive behavior supports (SWPBS). For example, the *Self-Assessment Tool* (SET) is a 28-item observation and interview instrument (Sugai, Lewis-Palmer, Todd, & Horner, 1999) that is conducted at the school by a trained coach or researcher. The *Benchmarks of Quality* (BoQ; Kincaid, Childs, & George, 2005) measurement was designed to be a more practical instrument for schools to assess strengths and weaknesses in their behavioral intervention strategies. This 53-item instrument is implemented through a process where the coaches and leadership team members rate each item as 'in place,' 'needs improvement,' or 'not in place.' Through discussions, consensus is reached on the status of implementation (Cohen et al., 2007). The *Effective Behavior*

Support survey (EBS Survey; Sugai, Horner, & Todd, 2003) can be completed by the entire school staff. This assessment asks staff to rate current status and priorities for improvement across behavior support systems (i.e., school-wide discipline systems, non-classroom management systems, classroom management systems, and systems for individual students).

To assess the quality and quantity of implementation related to literacy components, the *Planning and Evaluation Tool for Effective Reading Supports – Revised* (PET-R) focuses on goals, materials, time allocation for instruction, and other factors to rate overall implementation of a school's K-3 reading program (Kame'enui & Simmons, 2003). Like the EBS, the PET-R asks respondents to rate the level of implementation (i.e., not in place, partially in place, or fully in place).

Rubrics such as the *Phases of Implementation* rubric (Vermont Department of Education, 2011) constitute another way to measure implementation. This rubric is designed to be completed by the behavior coach or by an observer in order to identify implementation levels (e.g., emerging, implementing, and sustaining). Implementation rubrics have also been found helpful in providing a data source for school leaders to reflect on school-wide implementation of tiered models and plan for professional development (Liu, Alonzo, & Tindal, 2011). In general, the purpose of an implementation rubric is to outline operational definitions for implementation and provide illustration of best practices. Increasingly, application of rubrics for guiding reflection on instructional practices and school improvement is gaining attention.

Within tiered models, all of the measures described are designed to inform action planning and continued implementation of the model. Much is still unknown regarding treatment integrity or fidelity of implementation within tiered models, including the amount of allowable variation within implementation that can produce positive results for students. Additionally, it is important to understand which treatment components were not implemented in schools that failed to see growth in student academic or behavioral achievement (Bradshaw et al., 2008).

Rationale for a New Measure

Within integrated academic/behavior RTI models, there are clear gaps in available fidelity measures, and no existing measure evaluates implementation of the core features of integrated models from a whole school perspective in a cost-effective, minimally intrusive manner. Furthermore, measures that do exist are intervention-specific and not appropriate for integrated models. Without treatment integrity data, we have no way of understanding the variance in school gains. Finally, data produced by the existing measures often lacks a utilization focus, meaning that the data is not translated and visually represented in meaningful, easily-understandable way for continual planning and improvement.

Existing fidelity measures focus primarily on the perceptions of the small group of individuals that constitute the school leadership team. Consistency in implementation throughout the school is an assumption often made by RTI initiatives. However, assessment of treatment integrity is necessary to identify acceptable variations and interpret of outcomes (Center on Positive Behavioral Interventions and Supports, 2004; Cohen et al., 2007; Griffiths, Parson, Burns, VanDerHeyden, & Tilly, 2007; Noell & Gansle, 2006). Fuchs and Deshler (2007) provide a clear explanation:

"In reality, there are many situational factors – inside and outside the classroom – that support and account for [the RTI's] successful implementation. These factors are as important to identify and understand as are the components of the RTI model itself" (p. 131). School leadership team members do not necessarily represent the views of all school staff. In fact, Walker (2006) found that the implementation ratings of school leadership team members were significantly higher than those of other school staff.

Many current fidelity measures are time-intensive. Sight visits and classroom walk-throughs have been shown to take large amounts of evaluator, staff, and administrative time (Cohen et al., 2007). For example, the SET is often cited as a widely used and adapted measure of SWPBS (Horner, Todd, Lewis-Palmer, et al., 2004), but it requires four to six hours of an evaluator's time (e.g., travel, interviewing, observing, and scoring), as well as access to students, staff, and administrators (Cohen et al., 2007).

Furthermore, academic and behavior RTI models include unique process, fidelity, and outcome measures particular to each model. Measures that are intervention-specific limit their utility across interventions (Sanetti & Kratochwill, 2009), and it becomes overwhelming and time-intensive for school teams to implement numerous measures for multiple RTI models (Pool, Johnson, & Carter, 2010). As stated by the Florida Department of Education (2008):

> Clearly, each set of [RTI] efforts is built upon common elements, but with single-purpose resources and in segregated activities. Each separate effort also involves a unique set of terminology, professional development requirements, and data collection and reporting systems, which result in district and school personnel perceiving that an overwhelming number of parallel initiatives are either required or encouraged (p. 2).

While academic and behavioral outcomes vary across schools, with some schools showing higher gains than others, these variations cannot be fully understood or explained without first considering the implementation fidelity of the intervention (Griffith et al., 2007; Lane, 2007). There is a lack of such measures within combined academic and behavior RTI models, leading to a lack of fidelity evidence. Until new measures grounded in integrated academic/behavior RTI research are developed, this issue will continue, potentially leading to a lack of confidence in the success of tiered interventions.

Ensuring that data is both valuable for both schools and external evaluators/funders and also cost-effective is often a difficult balance. Utilization-focused evaluation directly addresses this balancing act. As Michael Patton (2008) summarized:

> *Utilization-focused evaluation* is evaluation done for and with specific intended primary users for specific, intended uses. *Utilization-focused evaluation* begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything done, from beginning to end, will affect use (p. 37, italics in original).

In other words, this evaluation approach supplies school decision-makers with high quality data to guide their actions toward systemic educational improvement. In addition to being valid and reliable, Holden, Friedman, and Santiago (2001) stated that evaluation "results must be intelligible and useful to multiple stakeholders" (p. 6). The need for data to be useful for school teams provides another layer of complexity to measuring fidelity of implementation.

Because the measures within academic and behavior RTI are specific to each model, integrated multi-tiered processes have identified the need for measures that address academics and behavior as well as school-wide implementation of essential elements of effective school systems. To address the needs of schools implementing an integrated academic/behavior RTI model and provide meaningful treatment integrity data in a cost-effective manner, a team of researchers and practitioners collaborated to develop the *School Implementation Scale*, a cost-effective, online survey instrument to be completed by all school personnel.

Methodology

Within the development and implementation of the *School Implementation Scale*, research questions guided the design. These questions included:

1. Can the measure provide reliable data across all school staff and rade levels?
2. Do the perceptions of school leadership team members accurately represent the perceptions of all school staff?
3. Are all components of tiered support implemented simultaneously?
4. Do staff perceptions of implementation correlate with student achievement?
5. Can the measure assess implementation across multiple tiered support models?

To address each question, the *School Implementation Scale* was developed and tested through an iterative design process. It was then administered across multiple years with a cohort of 14 Midwestern schools that were implementing an integrated academic and behavior RTI model.

Development of the School Implementation Scale

The *School Implementation Scale* was developed through a multi-year iterative design process that included: (a) framework conceptualization, (b) item development, (c) pilot testing, (d) item refinement, and (e) full implementation. Its design was conceptualized as a measure of school-wide implementation for an integrated academic/behavior RTI model of academic and behavior support being implemented in one Midwestern state. The components of this integrated model were based on a comprehensive review of essential elements of effective school systems identified by the National Center on Response to Intervention, the Technical Assistance Center on Positive Behavioral Interventions and Supports, Professional Learning Communities, Southern Regional Education Board High Schools that Work, and the U.S. Department of Education Reading First initiative (Jenson, 2008). As a result of Jenson's (2008) analysis, the essential elements of effective school systems (e.g., staff-

level intervention components and situational factors that support student success) of multi-tiered models were defined as:

1. *School Culture:* Joint commitment to a *shared vision* for success with *leadership* and a *collaborative environment* that supports and encourages continuous discussion, sharing, reflection, and problem solving toward a common goal.
2. *Ongoing professional development:* Research-based professional learning that is ongoing, tailored to the needs of participants, integrated within school improvement planning, and focused on student improvement. This includes *mentoring and coaching* to teach and support educators during implementation.
3. *Evidence-based practices*: Implementation of practices supported with empirical evidence of effectiveness using methods that are appropriate for the environment and circumstances. This includes *data-based decision making*, *universal screening* and *progress monitoring* (i.e., systematically using data from assessments to drive decisions for educational practices and target interventions) and *maximizing resources* to implement assessments and effective instructional practices.
4. *Family engagement*: Involvement of family members to inform and reinforce school improvement efforts and promote *culturally responsive practices*.

These essential elements are designed to be integrated into the school climate to drive decision making, facilitate innovation, and support student progress.

The evaluation team, which included university evaluators, state-level administrators, regional professional development providers, implementation coaches, and local-level educators, identified criteria for the development of the measure. First and foremost, the measure needed to be cost-effective and efficient to administer. Second, the survey needed to obtain perceptions from all school staff, not just from those on the school leadership teams. Third, the survey needed to provide results to schools in a format that was easy to interpret in order to inform team-level data-based decision making and action planning. To address these requirements, it was decided that an online survey would be developed with the evaluators of the project analyzing data to create school-level summary reports.

Item development for the *School Implementation Scale* began with a literature review of available school staff surveys within multi-tiered models, though a paucity of such surveys was found. In addition to the surveys identified in the previous section, the *Standards Assessment Inventory* developed by the National Staff Development Council addressed some components of multi-tiered models, specifically professional development, and was designed for school staff to complete.

To address the inflation in scores of self-report data (Welkenhuysen-Gybels, Billiet, & Cambre, 2003), items were developed to address individual implementation as opposed to perceived school-level implementation. For example, one item stated, "I regularly communicate with families regarding student academic & behavioral goals/progress." This was similar to the item in the Standards Assessment Instrument that states, "Teachers are provided with opportunities to learn how to involve families in their children's education," but the *School Implementation Scale* item required

identification of personal practices as opposed to perceptions of training that influence overall school practices. Because all staff completed the survey, composite school results provided whole-school implementation data. Additionally, school staff were trained on the essential elements of effective school systems, a strategy that has been to shown to increase the accuracy of self-assessment results (Irvine, 1983).

A survey of 99 items was developed, reviewed by experts of the integrated academic/behavior RTI model, and revised based on their feedback. In the spring of 2009, the survey was administered to all staff in six of the schools that were implementing the integrated model. School staff rated each item on a 5-point Likert scale, and summary reports were developed and provided to each school. Each school's action plan was then reviewed to identify how survey results influenced needs and strategy identification. In addition, the implementation coach for each school was asked the following questions: (1) "Was the data beneficial to the school leadership team?" (2) "Did the data assist the team in completing the action plan?" and (3) "What ideas did the school leadership team identify for improving the survey?" Reliability and factor analyses were conducted and items were removed from the survey based on these analyses. With 99 items, this survey was considered too time-intensive for continued school-wide implementation.

In the fall of 2009, a team of state, regional, and local professional development providers with experience implementing tiered reform models reviewed the survey and participated in multiple focus groups discussing further revision of the instrument. An updated survey consisting of 31 items was piloted in the spring of 2010 in the six schools from the first pilot and an additional eight schools that were implementing the integrated multi-tiered model. Once again, reliability and factor analyses informed further item revision and the addition of two items to the survey. This process provided preliminary evidence of construct validity (Messick, 1988).

This final measure was implemented across the same 14 schools during the 2010-11 school year and again in the 2011-12 school year. It was also administered in an additional 11 schools that were in beginning stages of implementation of the integrated multi-tiered model and in three schools implementing a separate integrated model. The results identified in this study provide the data analyses and interpretation of this fully-developed measure of school staff implementation in integrated academic/behavior RTI models.

Administration and Scoring

The *School Implementation Scale* was administered within each school through a standardized process. This process included providing written justification for administration and instructions to each school's principal, with sample text for e-mail dissemination. The principal then sent an email to all school staff requesting completion by a specific date. It was recommended that a two-week window be provided for completion. The written instructions also suggested that the principal follow-up with a reminder email toward the end of the administration window.

School staff clicked on the link within the email provided by their principal. This link took them directly to the survey that included 33 Likert-scale items and 3 demographic items (i.e., school, professional role, and membership in

the school leadership team). Completion of the entire survey was estimated to take 5 to 10 minutes, and responses were automatically saved by the online system.

At the end of the administration timeframe, the evaluators developed a summary report for the school that provided a table reporting scores and means for each item and domain, as well as stacked bar graphs for the same data. Development of the summary reports required 30 minutes of evaluator time for each school. It was recommended to administrators and school leadership teams to share the visual displays of the data with all school staff and to use the results within their data-based needs assessment and action planning process.

Settings and Participants

Between January and March, 2011, 346 participants from eleven public elementary schools and three public middle schools completed the *School Implementation Scale*. Each of these schools was in their second year of implementation of an integrated multi-tiered support model. Using the National Center for Education Statistics (2011) urbanicity classifications, these schools represented two city, two suburban, three town, and six rural districts from across all geographic regions of a Midwestern state. Free and reduced lunch rates ranged from 21% to 77% of the school population, and the percentage of students from ethnically diverse groups ranged from 1% to 95%. Total teachers within the schools ranged from nine to fifty-four. School staff were categorized as teachers (294), administrators (15), other certified staff (24), and noncertified staff (11).

To provide additional yearly and grade-level comparisons, data from an additional nine schools within the districts of the 14 original schools were included in the analyses. These schools, which were in their first year of implementation, included three elementary, one middle, and four high schools.

Three urban charter schools in another state implementing a different integrated academic/behavior RTI model also completed the *School Implementation Scale* during the 2010-11 school year. These schools included one elementary, one middle, and one K-8 school. The data from these schools were used to analyze the reliability of the scale for a separate RTI model.

Analysis

A variety of quantitative analyses were conducted to respond to the research questions. These included reliability estimates, exploratory factor analyses, *t*-tests, and correlations. Together, the results of these analyses provided a comprehensive description of the reliability of the *School Implementation Scale* and initial outcomes related to implementation of the integrated academic/behavior RTI model.

Validity is an important but broad construct that encompasses both technical components and social ramifications (Messick, 1988, 1994). Validity includes concurrent, predictive, and construct validity as well as the use of results (Gersten, Keating, & Irvin, 1995). Within the development and analysis of the *School Implementation Scale*, validity considerations included technical aspects, patterns of ratings across roles and grade levels, trends across subdomains, and use of results.

Reliability refers to the consistency or reproducibility of results over repeated administration and across items. A reliable survey is as free as possible from measurement error generated from poorly worded questions, poor design, ambiguous terms, inappropriate reading level, or unclear directions. Cronbach's coefficient alpha was used to evaluate the reliability because it could be used to analyze a single form survey and could be applied to ordinal data (Gliem & Gliem, 2003). This method identified the consistency of responses across items or subsets of items. In addition, exploratory factor analysis (EFA) was performed to test the content homogeneity, which, in turn, served as evidence to the reliability of the survey (DeCoster, J., 1998; Green & Salkind, 2011).

Comparisons among groups and across years were evaluated through *t*-tests and analysis of variance (ANOVA) statistical methods. An *a priori* significance level was set at .05 for all analyses. Additionally, effect size estimates were reported for each analysis (Lowry, 2011; Green & Salkind, 2011).

Results

The results of this study revealed that the *School Implementation Scale* is a highly reliable instrument that provides valid data on the implementation of integrated academic/behavior RTI models within schools. In addition, initial findings reveal that the *School Implementation Scale* has the potential to evaluate school-wide practices that correlate to higher levels of student proficiency on state communication arts assessments.

Reliability of the School Implementation Scale

Items within the *School Implementation Scale* were intentionally designed to measure implementation of each of the essential elements of effective school systems. The exploratory factor analysis revealed that the items also group within three implementation levels: school-level (i.e., systems components); classroom-level (i.e., instructional components), and individual student-level (i.e., individualized support components). Table 1 provides the overall Coefficient alpha, subdomains by essential element, and level of implementation.

Reliability across roles. School staff consisted of teachers, administrators, other certified staff, and noncertified staff. Reliability estimates for all roles were above .90 (see Table 2). This reveals a high internal consistency for different subgroups of school staff. In other words, regardless of school role, each respondent was consistent across items. The survey measured the same construct in essentially the same way and had high content homogeneity and item quality.

Reliability was not impacted by involvement on the school leadership team (i.e., the team of 5-15 school-level staff that obtained additional training in the model and worked collaboratively to implement the model components within the school). Full-scale reliability estimates for team members and non-team members were .959 and .969 respectively.

Table 1. ***Reliability by Essential Element and Level of Implementation*** **(*n* =346)**

Essential Element	Number of Items	Co-Efficient Alpha
School Culture	10	.888
Ongoing Professional Development	6	.866
Evidence-Based Practices	12	.902
Family Engagement	5	.808
Level of Implementation		
School-Level Implementation	11	.906
Classroom-Level Implementation	11	.898
Student-Level Implementation	11	.913
Total Instrument	33	.961

Table 2. ***Reliability by Role of Respondent***

	Teachers	Administrators	Other Certified Staff	Noncertified Staff
Respondents	294	15	24	11
Alpha	.960	.929	.953	.970

Reliability across grade levels. To test the reliability of the *School Implementation Scale* across grade levels, the survey was administered between September and November 2011, in 14 elementary, four middle, and five high schools. Results revealed consistently high reliability across school levels (see Table 3). In other words, the *School Implementation Scale* has the potential to provide consistent data across all grade levels.

Table 3. ***Reliability by School Level of Respondent***

	Elementary	Middle	High
Respondents	343	128	82
Alpha	.955	.929	.968

Reliability within another RTI model. A final test of reliability was conducted with three urban charter schools (one elementary, one middle, and one K-8) in another state implementing a separate integrated academic/behavior RTI model. A total of 36 staff from these schools completed the *School Implementation Scale,* producing a full-scale coefficient alpha of .907. Subscale alphas by level of implementation were all above .850 and alphas by essential element fell between .753 and .782. This analysis provided supporting evidence that the *School Implementation Scale* could produce reliable results across RTI models. Further analysis across other RTI models and with additional schools will be necessary to confirm the use of the scale across models.

Outcomes

The results of the *School Implementation Scale* were used to identify variations across groups, grade levels, and subdomains. In addition, analyses were conducted to identify outcomes related to school-wide implementation and student academic achievement.

Variation among groups. As identified in the literature review, most measures of implementation within multi-tiered models have been designed to be completed by the school leadership teams. To identify the effectiveness of this team in representing the perceptions of all school staff, independent sample *t*-tests were conducted. The test was significant, with school leadership team members consistently rating implementation higher than other school staff (see Table 4). While both groups provided consistent responses, the members of the school leadership team perceived higher levels of implementation than full-school data supported.

As shown by the results, it is necessary to obtain data from school leadership team members as well as from other staff members. The eta square index indicated that 2% of the variance in mean scores was accounted for by team membership. While there was a small effect size, the results indicated that members of the school leadership team cannot accurately represent the perceptions of all school staff. This supports the findings of Walker (2006) showing that participation in a leadership team influenced ratings of implementation of a school-wide positive behavioral supports model.

Table 4. ***Group Membership Impact on Mean Scores***

	Team Member (*n*=104)		Not Team Member (*n*=237)		
Scale	*Mean*	*SD*	*Mean*	*SD*	*Sig*
Full Scale	4.10	.62	3.89	.70	<.01
School Culture	4.21	.62	3.95	.74	.08
Ongoing Professional Development	4.13	.73	3.90	.80	.02
Evidence-Based Practices	4.00	.69	3.80	.75	.02
Family Engagement	4.10	.63	3.96	.72	.20

Variation across subdomains. Systems-level implementation was reported to be in place at higher levels than classroom and individual student implementation. A one-way ANOVA contrasted the mean scores across the three levels of implementation (i.e., school, classroom, and individual student). Results indicated significant differences across the three subscales, $F(2, 1035) = 6.39$, $p = <.01$, partial eta squared .012. Follow-up Tukey tests revealed that significant differences existed between the student-level subscale ($M = 3.84$; $SD = .77$) and both of the other subscales: school-level ($M = 4.03$; $SD = .71$; $p = .01$) and classroom-level ($M = 3.99$; $SD = .70$; $p = .023$). These results support Safran's (2006) finding that individual student-level components were judged to be the least in place, with lower mean scores than classroom-level and school-level domains when initially implementing an RTI model.

When comparing the essential elements of effective school systems, results indicated differences across the four subscales, $F(3, 1380) = 3.46$, $p = .016$, partial eta squared .007. Follow-up Tukey tests revealed that significant differences existed between the school culture subscale ($M = 4.03$; $SD = .71$) and the evidence-based practices subscale ($M = 3.86$; $SD = .73$; $p = .015$). Together, these analyses indicated that individual student-level and evidence-based practices components of the integrated academic/behavior RTI model were the least in place among the 14 schools. These results suggest that RTI implementation began at the school level by developing a positive school culture with a shared vision among staff with strong and supportive leadership.

Implementation across years. The fully-developed *School Implementation Scale* was administered across two years in 23 schools (14 elementary, four middle, and four high schools). Significant increases in the implementation of essential elements of effective school systems were identified, including full-scale, subdomains of school culture, ongoing professional development, and family engagement as well as the student and school levels of implementation. Because the survey respondents were anonymous, paired samples *t*-tests were not appropriate and therefore independent sample *t*-tests were conducted. As indicated in Table 5, implementation of the integrated academic/behavior RTI model increased across the two years of implementation.

To assess the degree of implementation within each school, participant ratings were converted to binomial scores. Responses of 'true of me now' (4) and 'very true of me now' (5) were coded as implementation, while responses of 'somewhat true of me now' (3) and below were coded as lack of implementation. The threshold of 80% of school staff implementing each multi-tiered component was set as the definition of school-wide implementation. This threshold was consistent with other school-wide implementation measures (e.g., *Self-Assessment Tool*, Sugai et al., 1999) as the level of sufficient full-school implementation for staff buy-in of the model (Simonsen, Sugai, & Negron, 2008). For the 14 schools in their second year of implementation during the 2010-11 school year, five of the 33 survey items were found to be fully implemented (two within the school-level domain and three within the classroom-level domain). During the 2011-12 school year, all items that met the school-wide implementation threshold the previous year continued to be implemented by 80-100% of school staff, and an additional seven items reached the threshold. Together, these fully-implemented practices included four items in the school-level do-

main, five items in the classroom-level domain, and three items in the student-level domain. When analyzed by essential element, these items represent practices related to each component including school culture (5), ongoing professional development (1), evidence-based practices (4) and family engagement (2). Overall, a higher percentage of school staff were implementing components of the integrated academic/behavior RTI model as implementation continued across multiple years.

Table 5. *Implementation across Years*

Scale	Year	*N*	*M*	*SD*	*Sig*	*d*
Full Scale	2010-11	784	3.81	.72	.02	.12
	2011-12	696	3.89	.67		
School Culture	2010-11	784	3.89	.74	.02	.12
	2011-12	696	3.97	.69		
Ongoing Professional Development	2010-11	780	3.79	.85	.02	.12
	2011-12	694	3.89	.76		
Evidence-Based Practices	2010-11	781	3.71	.81	.08	.09
	2011-12	691	3.78	.77		
Family Engagement	2010-11	766	3.90	.72	.01	.14
	2011-12	674	4.00	.68		
Level of Implementation						
School-Level Implementation	2010-11	784	3.89	.76	.01	.13
	2011-12	696	3.98	.68		
Classroom-Level Implementation	2010-11	778	3.82	.77	.07	.09
	2011-12	693	3.89	.75		
Student-Level Implementation	2010-11	780	3.69	.82	.03	.11
	2011-12	691	3.78	.79		

Correlation with academic achievement. Using the degree of implementation method identified above (i.e., percentage of respondents that rated each item as a 4 or 5), school-wide implementation was identified. For the 14 schools in their second year of implementation of the integrated academic/behavior RTI model, the percentage of the 33 *School Implementation Scale* items with 80-100% of staff rating the item as a 4-5 ranged from 1 item (3%) to 30 items (91%). This showed that there was great variation in the implementation across the 14 schools. There was also a high level of variation among the schools in the communication arts state assessment scores for students with disabilities. Between 2009 and 2011, the change in the percentage of students with disabilities that met proficiency on the state communication arts assessment ranged from -17.5% to +43.3% among these schools.

The percentage of items rated 4-5 by 80-100% of staff within each school was compared to the increase in the percent of students with disabilities that met

proficiency on the state communication arts assessment. Results revealed a moderate correlation coefficient (0.55) between the two. In other words, schools that implemented the essential elements of RTI school-wide were more likely to see increases in reading and writing achievement of students with disabilities. This supported recent research that identified the relationship between higher treatment integrity and improved student outcomes in academics and behavior (Biggs, Vernberg, Twemlow, Fonagy, & Dill, 2008; Childs et al., 2010).

Limitations

The *School Implementation Scale* is a 33-item self-report measure. As with all self-report measures, a primary limitation is acquiescent response bias (Welkenhuysen-Gybels et al., 2003). To mitigate this limitation, school administrators were provided with standard instructions for requesting survey completion. These instructions included the request to answer honestly so that results could inform continued improvement through action planning and assurance that all responses would be anonymous. Additionally, survey response options were purposely constructed to ask school staff to only report on themselves (e.g., 'this is very true of me now'). Survey items were also designed to ask about individual practices (e.g., I receive coaching/mentoring to help me implement evidence-based instructional practices), not perceptions of school-wide practices. Additionally, through in-service training and coaching within each school as part of the integrated model implementation, staff were trained in data-based decision making and self-assessment behaviors. This training has been shown to increase the accuracy of self-rating measures (Irvine, 1983). Overall, Likert scales have been found to provide reliable and valid information (Maurer & Anders, 2000).

Another limitation of self-report scales is the response rate and characteristics of respondents. Because the *School Implementation Scale* was administered anonymously, respondents only were required to identify their school and role, and therefore it was not possible to identify the specific non-respondents. To mitigate response rate limitations, school administrators sent the survey to all school staff requesting completion by a specific date. In most cases, these same administrators provided reminders to staff, either via email or face-to-face communication. During the 2010-11 school year, the response rate of the 14 schools ranged from 74% to 100%, with an overall response rate of 95%.

Discussion

Integrated academic/behavior RTI models are being widely developed and implemented by schools, and there is a need for cost-effective, evidence-based treatment integrity measures to better understand the impact of these models. Evaluators and school teams need access to easy-to-use measures that incorporate the essential elements of effective school systems instead of multiple model-specific redundant measures and processes. Just like the RTI models, measures must also be integrated to ensure that they have the potential to inform school-wide data-based decision making, evaluate the implementation and effectiveness of the RTI model, and compare implementation components across multiple models. This study evaluated the salient empirical properties and outcomes of the *School Implementation Scale* within one

integrated academic/behavior RTI model. With the discussion, we report practical applications of the *School Implementation Scale* and implications for future research.

Practical Applications

As identified in the results section, the *School Implementation Scale* is a highly reliable instrument that demonstrates potential to evaluate the treatment integrity of integrated academic/behavior RTI models in schools. Analyses showed a correlation between school staff implementation of essential elements of effective school systems and an increase in reading and writing achievement for students with disabilities. Resulting data has been utilized by both school and state teams for ongoing planning, refinement and improvement in the integrated academic/behavior RTI model.

As discussed earlier, utilization-focused evaluation provides a framework for ensuring that evaluation results are meaningful to stakeholders and can be incorporated into data-based decision making (Patton, 2008). Together with the development and revision of the *School Implementation Scale*, school summary reports also followed a similar iterative process. After the first administration of the scale, results were provided to schools as mean scores by domain and stacked bar graphs for each item. While this data was easy to view, it lacked the in-depth results requested by schools. To increase usefulness of the data, tables identifying the number of respondents at each rating level were added. The mode response was highlighted to provide a visual representation of the response option with the highest percentage of respondents. School administrators reported that this data provided the information necessary for annual action planning, but as they accumulated multiple years of data, making comparisons across the years became a higher priority.

For each school, annual comparison reports were developed to illustrate change in level of implementation of essential elements of effective school systems across school years. Originally, this data was reported as mean scores with bar graphs, but further analysis revealed that data related to attainment of school-wide implementation (i.e., 80% "True of Me Now" or "Very True of Me Now" responses) could be more beneficial. This reporting format also proved to be valuable to external evaluators of the project in assessing the degree of implementation of the model.

In addition to school utilization of data, the state leadership team reviewed summarized *School Implementation Scale* data a minimum of twice per year in order to provide targeted professional development. Data revealed specific areas of need for participating schools, and professional development in these areas was offered through trainings held three times per year. For example, in the first year, data revealed a need for support in evidence-based practices, particularly those related to progress monitoring. State leadership collaborated with two regional professional development centers to offer ongoing professional development in beginning and intermediate data processes. Furthermore, the state included a pre-post knowledge survey in the trainings and identified a significant increase in understanding of progress monitoring components. The state leadership team has also used summary data over time to illustrate how pilot schools increased adoption of essential elements of effective school systems and the impact of this adoption on academic gains. This information was included in a concise annual report for statewide dissemination to increase awareness of the model and recruit additional schools.

Future Research

School improvement efforts have entered an exciting time with the acknowledgment that integrated academic/behavior RTI models can have a positive impact on both students with and without disabilities to achieve academically (Missouri Department of Education, 2011). Substantial research is required to ensure that these tiered support models continue through a self-correcting feedback cycle informed by data (Fuchs & Deshler, 2007; Kansas Department of Education, 2011). Additionally, as more RTI models are developed and states expand the implementation of their customized models, a common lexicon is vital for discussion among researchers; administrators at the state, regional, and local levels; and practitioners. By focusing on the essential elements of effective school systems, conversations can form around a common language free from initiative-specific jargon. This, in turn, can advance the field of school-wide educational reform and evaluation. Effective strategies for implementing integrated academic/behavior RTI models should also be made available across initiatives. By organizing and reporting these strategies by the essential elements (i.e., school culture, ongoing professional development, evidence-based practices, and family engagement), strategies can be shared across initiatives in a meaningful way.

Along with a common language and effective strategies, it is necessary to develop common evaluation measures that can inform data-based decision making within schools, evaluate the outcomes of multiple models, and possibly even compare models to provide a framework for continued improvement and facilitate the cross-flow of information related to effective practices.

Specifically related to the *School Implementation Scale*, it is expected that future research will include the implementation of the scale within schools implementing other RTI models. To this end, steps are underway to make the survey available online to schools across the country. Along with the expansion of implementation, continued administration of the scale across multiple years will build the research base of survey results. Further evaluation of the utility of the *School Implementation Scale* could include interviews and focus groups with school and district leadership to determine their perceptions regarding the ease of administration, analysis of the data, sharing of the results with school staff, and strategies to increase implementation of essential elements of effective school RTI systems.

References

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

Barton-Arwood, S., Morrow, L., Lane, K., & Jolivette, K. (2005). Project IMPROVE: Improving teachers' ability to address students' social needs. *Education and Treatment of Children, 28,* 430-443.

Biggs, B.K., Vernberg, E.M., Twemlow, S.W., Fonagy, P., & Dill, E.J. (2008). Teacher adherence and its relation to teacher attitudes and student outcomes in an elementary school-based violence prevention program. *School Psychology Review,* 533-549.

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,* 1-26.

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

Center on Positive Behavioral Interventions and Supports. (2004). *School-wide positive behavior support: Implementers' blueprint and self-assessment.* Eugene, OR: University of Oregon.

Childs, K.E., Kincaid, D., & George, H.P. (2010). A model for statewide evaluation of a universal positive behavior support initiative. *Journal of Positive Behavior Interventions, 12,* 198-210.

Cohen, R., Kincaid, D., Childs, K. (2007). Measuring school-side positive behavior support implementation: Development and validation of the Benchmarks of Quality. *Journal of Positive Behavior Interventions, 9,* 203-213.

DeCoster, J. (1998). *Overview of factor analysis.* Tuscaloosa, AL: University of Alabama. Retrieved from http://www.stat-help.com/factor.pdf.

Florida Department of Education. (2008). *Statewide response to instruction/intervention implementation plan.* Tallahassee, FL: Author.

Florida's Positive Behavior Support Project. (2011). *Implementing a multi-tiered system of support for behavior: A practical guide.* Tampa, FL: Author.

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

Gersten, R., Keating, T., & Irvin, L.K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children, 61,* 510-519.

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

Goodman, S. (2006). MiBLSi outlines a multi-level system of support for Michigan students. *Focus on Results, 7*(2), 1-4.

Gresham, F.M., MacMillan, D.L., Beebe-Frankenberger, M.E., & Bocian, K.M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*, 198–205.

Green, S.B., & Salkind, N.J. (2011). *Using SPSS for Windows and Macintoch: Analyzing and understanding data* [6th Edition]. Upper Saddle River, NJ: Prentice Hall.

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

Holden, E.W., Friedman, R.M., & Santiago, R.L. (2001). Overview of the national evaluation of the Comprehensive Community Mental Health Services for Children and Their Families program. *Journal of Emotional and Behavioral Disorders, 9*(1), 4-12.

Horner, R.H., Todd, A., Lewis-Palmer, T., Irvin, L., Sugai, G., & Boland, J. (2004). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Intervention 6*, 3-12.

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

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

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

Kame'enui, E.J., & Simmons, D.C. (2003). *Planning and Evaluation Tool for Effective Schoolwide Reading Programs – Revised (PET-R).* Eugene, OR: University of Oregon Center on Teaching and Learning.

Kansas State Department of Education. (2011). *Kansas Multi-Tier System of Supports: Structuring guide.* Topeka, KS: Kansas MTSS Project, Kansas Technical Assistance System Network.

Kincaid, D., Childs, K., & George, H. (2005). *School-Wide Benchmarks of Quality - Revised* (Unpublished instrument). University of South Florida, Tampa.

Lane, K.L. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children, 30,* 135-164.

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

Liu, K., Alonzo, J., & Tindal, G. (2011). Implementation and outcomes. In E. S. Shapiro, N. Zigmond, T. Wallace & D. Marston (Eds.), *Models of Implementing Response to Intervention: Tools, outcomes, and implications* (pp. 341-374). New York, NY: Guilford Press.

Lowry, R. (2011). *Concepts and applications of inferential statistics.* Poughkeepsie, NY: Vassar College. Retrieved from http://faculty.vassar.edu/lowry/webtext.html.

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

Maurer, T., & Andrews, K. (2000). *Traditional, Likert and simplified measures of self-efficacy. Educational and Psychological Measurement*, 60, 965-973.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 33-48). Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 12-23.

Missouri Department of Education. (2010). *Missouri Integrated Model: Annual dissemination report.* Jefferson City, MO: Author.

Missouri Department of Education. (2011). *Missouri Integrated Model: Annual dissemination report.* Jefferson City, MO: Author.

Noell, G.H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W.P. Erchul & S.M. Sheridan (Eds.), *Handbook of research in school consultation: Empirical foundations for the field* (pp. 315-334). Mahway, NJ: Erlbaum.

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

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Pool, J.L., Johnson, E.S., & Carter, D.R. (2010). *Implementing a combined RTI/PBS model: Social validity*. Retrieved from http://www.rtinetwork.org/rti-blog/entry/1/97.

Sanetti, L.M., & Kratochwill, T.R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445-459.

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

Spaulding, S.A., Horner, R.H., May, S.L., & Vincent, C.G. (2008, November). *Evaluation brief: Implementation of school-wide PBS across the United States.* OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. Retrieved from http://pbis.org/evaluation/evaluation_briefs/default.aspx

Sugai, G., & Horner, R.H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24,* 23-50.

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

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

Sugai, G., Horner, R.H., Lewis-Palmer, T. (2001). *Team Implementation Checklist (TIC).* Eugene, OR: University of Oregon.

Vermont Department of Education. (2011). *Vermont PBIS – Phases of Implementation (PoI) Rubric for annual celebration.* Montpelier, VT: Author.

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

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