

Validation Study of the Implementation Rubric for RTI in Reading (I-RRR) at Elementary Schools

Kimy Liu^{*}, Jacob Werblow^{**}

Abstract

Response-to-Intervention (RTI) is a complex process that addresses how to diagnose, label, instruct and evaluate low-performing students. Despite RTI being a common school practice in public elementary schools across the United States, there is little consensus on how to implement RTI within educational settings. In this validation study, the authors describe the development of a rubric that aims to evaluate the implementation of RTI as a school-wide reading intervention in elementary schools. Using a test development procedure (Downing & Haladyna; 2006), the rubric was developed based upon a review of the literature, validation by known experts in the field, and a comparison by two independent raters to evaluate two schools' RTI implementation. Through this process, six components were identified, which include evaluating: Tier 1 instruction, universal screening procedures, Tier 2 and Tier 3 interventions, progress monitoring practices, evidence-based decision-making, and organizational supports. Results suggest that the Implementation Rubric for RTI Assessment in Reading (I-RRR) is a valid and reliable measure of the effectiveness of RTI implementation.

Keywords: RTI, reading, rubrics, validation, elementary schools.

* California State University, Stanislaus

** Central Connecticut State University

Introduction

Half of all students with disabilities in America receive special education services under the category of specific learning disabilities. In the past, educators used a “discrepancy model” to qualify students under SLD if they had unexplained discrepancy between their academic performance and their potential (as evidence by IQs and their achievement). For years, researchers and practitioners had been unsatisfied with the discrepancy model because it often withheld much-needed early interventions and services to students who were on the verge of academic failure. Over the last two decades, Response to Intervention (RTI) has evolved to rectify the flaws of identification process anchored on discrepancy model.

Since 2008, nearly every state department of education in the country, including the District of Columbia, has practiced or implemented RTI (Hoover, Baca, Wexler-Love, Saenz, 2008). Although this approach has become a widespread school practice in recent years, there is wide variation in the ways that RTI is implemented within and among school districts across the country (Berkeley, Bender, Peaster, & Saunders, 2009; Hughes & Dexter, 2016). In the *Response to Intervention Adaption Survey: 2010* (spectrum k12, 2010) RTI for reading was adopted in 90% of the 1,101 elementary schools sampled. Among schools with at least some level of RTI implementation, nearly half of the administrators reported that their district did not have any implementation plan. This points to the need for an objective measurement of successful RTI implementation within schools as educators need a valid a reliable way to evaluate the implementation of response to intervention (RTI).

Successful school reform must be anchored on a coherent logic model. How an RTI approach is enacted in a school system may well depend on the perspective of the key decision makers involved. While some design their RTI system as a problem-solving model (Lauet al., 2006; Marston, Lau, & Muskens, 2007; Peterson, Prasse, & Shinn, 2007; Vaughn & Denton 2008), its data-driven decision-making process begins with identifying the problems, planning the intervention, implementing the intervention, assessing the student learning outcomes, and using this assessment data to improve instructional design and delivery. Other RTI systems focus more on an approach that emphasizes a standard protocol, which was originally developed in public health service, and subsequently applied in the school settings, with the three-tiered model is its most popular variation, to promote school-wide positive behavior support (Fuchs & Deshler, 2007; Fuchs, Fuchs, & Compton, 2012; Hurlbut & Tunks, 2016; Shapiro, Zigmond, Wallace, & Marston, 2011; Stahl, Keane, & Simic, 2013). Another category of RTI systems use a hybrid of the two models in designing their RTI programs (Spencer, Petersen, & Adams, 2015; Van Der Heyden, 2011; Van Der Heyden, Witt, & Gilberton, 2007).

The misalignment between the intended goals and selected model can create challenges in implementation. If those designing the RTI process prefer a problem-solving model, the intervention provided will be more tailored to individual students' instructional needs. In contrast, those who focus on RTI from a standard protocol model emphasize adhering to standardized administration guidelines and proven effective intervention programs. The differences between these approaches have led to confusion over how schools are to implement the process of RTI (Burns, Appleton, & Stehouwer, 2005; Fuchs, et al., 2012; Reschly, Coolong-Chaffin, Christenson, & Gutkin, 2007). This confusion poses a challenge for schools. Teachers and administrators wish to implement the advocated Federal policy, yet they have no clear guidelines for determining whether or not they are implementing it appropriately. Because the goal of RTI is to coordinate the resources in the school system to improve student learning, practitioners need a method for evaluating the quality of RTI implementation at both the school and student levels.

In this study, the authors define RTI as a complex process that addresses how to diagnose, label, instruct and evaluate low-performing students. Under this definition, RTI is anchored upon a problem-solving cycle and/or multi-tiered service delivery approach (Denton, Vaughn, & Fletcher, 2003; Fuchs & Deschler, 2007; Fuchs, Mock, Morgan, & Young, 2003; Fuchs, 2003; Marston, Muyskens, Lau, & Canter, 2003). Given this classification, the authors present the development of a rubric designed to evaluate the quality of RTI implementation at its host environment.

Six Essential Components of RTI

The first step in developing a way to measure RTI effectiveness was to conduct a literature review to identify the constructs to be measured. A literature review identified 28 peer-reviewed articles that mentioned progress monitoring and Tier 1, Tier 2 and Tier 3 interventions, or three-tiered models or multi-tiered interventions. Ten components of RTI were initially identified from these articles and later refined to six distinctly different components. Each of these six identified components is briefly described in the following section, with an emphasis on how each is implemented in schools to influence students' learning outcomes.

Effective Tier 1 Instruction

Tier 1 Instruction is defined as the comprehensive core reading curriculum instruction provided to all students in general education classes (Fuchs et al., 2003; Graden, Stoller, & Poth, 2007; Marston, Pickart, Reschly, Heistad, Muyskens, & Tindal, 2007; McMaster & Wagner, 2007; Stahl, Keane, & Simic, 2013) for a sufficient

amount of time to meet grade-level instructional goals (Harn, Kame'enui, & Simmons, 2007; Justice, 2006; McMaster & Wagner, 2007). Teachers should teach all “five big ideas” specified in National Reading Panel Report (2000) (Foorman & Torgesen, 2001; O'Connor, Bocian, Beebe-Frankenberg, 2005), or relevant ones to specific grade-level (Good & Kaminski, 2003; Marston, 2005), using evidence-based strategies (National Reading Panel, 2000). Because the quality of Tier 1 instruction is the first step in identifying students at-risk for developing reading problems (Speece et al., 2003), the I-RRR should also include criteria that are designed for teachers to evaluate the provided instruction (e.g., content of instruction, instructional time, use of instructional time, student time on task, satisfaction).

Universal Screening

The purposes of universal screening are three-fold: to determine the quality of general education (Tier 1) instruction, to identify students who are at-risk for long-term difficulties in learning to read, and to cluster students to form homogenous intervention groups for targeted and focused differentiated instruction (Foorman & Ciancio, 2005; Kamps & Greenwood, 2003). Practitioners use universal screening to determine whether the instructional quality in general education classes is sufficient to bring most students to grade-level benchmarks and what progress rate can be reasonably expected of individual students (Compton et al., 2012; Foorman & Ciancio, 2005; Fuchs, 2003; Jenkins, 2003; Van Der Heyden et al., 2003). The selected articles were consistent in recommending that universal screening be administered to all students three times a year to measure the students' progress against given criteria, such as a grade-level benchmark standard (Jenkins, 2005; O'Connor, Fulmer, & Harty, 2003; Vaughn, 2003). Four types of screening measures were typically referenced in articles about RTI implementation: curriculum-based measurement (CBM), informal inventory of students' academic skills, high-stake state or district assessment and norm-referenced standardized achievement tests. To justify its proposed use, a universal screener must also have appropriate documentation and evidence of validity and reliability (Fuchs, et al., 2012; Stahl, Keane, & Simic, 2013; Van Der Heyden, 2011).

Effective Tier 2 and Tier 3 Interventions

Consistently across the different studies reviewed, Tier 2 and Tier 3 Interventions were identified as appropriate for students identified as at-risk of developing reading difficulty or who did not respond to Tier 1 Instruction, as indicated by their level and rate of progress being lower than that of their peers (Fuchs, 1995; Fuchs, 1998; Harnet et al., 2007; Kamps & Greenwood, 2003). Recent studies (Compton et al., 2012; Fuchs et al., 2012) suggest that students can be accurately identified as

needing Tier 3 supports using only Tier 1 data, thus students do not need to fail in Tier 2 before being placed in Tier 3. It should be also noted that Tier 2 and Tier 3 interventions differ from Tier 1 instruction in their targeted population, purpose, instructional emphases, frequency, duration and intensity of instruction (Harnet al., 2007; Marston, 2005; Mellard, McKnight, & Jordan, 2010; O'Connoret al., 2005; Vaughn et al., 2012), and must focus on foundational and prerequisite skills needed for the student to make progress (Gersten, et al., 2009). Specifically, Tier 3 interventions, distinguishable their level of intensity and frequency from those in tier 2, are necessary to prevent school failure for about 5% of the school population (Fuchs et al, 2008; O'Connor et al., 2010; Simmons et al., 2011; Vaughn et al, 2010). To ensure that Tier 2 and Tier 3 Interventions are implemented with fidelity, interventions should be monitored to ensure they are taught as they are intended (Fuchs, 1998; O'Connoret al., 2005; Torgesen et al., 2001).

Progress Monitoring

In the literature reviewed, progress monitoring is defined as the regular collection of students' responses to a chosen assessment. The purpose of progress monitoring is to document students' incremental change on the targeted early literacy skills (Good & Kaminski, 2003) and to gather evidence on whether students have responded to additional instructional supports provided within an RTI approach (Christ & Hintze, 2007; Justice, 2006; Marston et al., 2007b; Tindal, Yovanoff, & Alonzo, 2006; Van Der Heyden & Jimerson, 2005). Across the studies synthesized, students' classification as Tier 1, Tier 2 or Tier 3 changes as a function of instruction and learning (Fletcher et al., 2005). In the I-RRR, the implementation of progress monitoring is judged by the technical adequacy of the measures, the quality of data collection, and the presentation of data (Good et al., 2002; Good & Kaminski, 2003). Without evidence of technical adequacy, the use of progress monitoring measures might not be justifiable.

Evidence-Based Decision-Making

Evidence-based decision-making is the logic and rationale behind RTI (Reschly et al., 2007). In the articles reviewed, evidence-based decision-making is often mentioned in the context of progress monitoring and consulting with a multi-discipline team (i.e., special and general education teachers, specialists, administrators) to evaluate how well students were responding to instruction (Bollman et al., 2007; Burns, Appleton, & Stehouwer, 2005; Fuchs, et al., 2012; Stahl, Keane, & Simic, 2013). In this study, the implementation of evidence-based decision-making is judged by what data are used to make the decisions, how the decisions are made, and what decisions are being made.

Organizational Support

Successful RTI implementation requires the collaboration of school faculty and personnel, who might not otherwise work together (Stahl, Keane, & Simic, 2013). This collaboration is important in designing and implementing research-based reading instruction and progress monitoring systems of RTI within a school (Grimes & Kurns, 2003; Kamps & Greenwood, 2003; Johnson et al., 2006; Marston, 2005). For the purpose of this study, organizational support is defined as strong leadership that organizes administrative support to provide effective professional training and encourage professional collaboration. This operational definition is distilled from studies mentioned above, including Little (1993), Darling-Hammond and McLaughlin (1995), and Desimone (2002). Principals are responsible for guiding the school team to set the goals and objectives, allocating resources, coordinating administrative supports to facilitate implementing research-based reading instruction and assessments, and monitoring the progress of program change.

Purpose and Goals

The purpose of the present study was to develop a valid, reliable, and useful rubric for evaluating the implementation of the RTI approach. Because the decision regarding students' responsiveness to instruction is determined at the student level, a useful RTI rubric must document and evaluate the instructional support and assessment provided to individual students. At the same time, RTI is consistently identified in the literature as a systems-wide approach to service delivery. Thus, to be most useful, an RTI rubric must also be able to capture systems-level information. In this document, we use the term: Implementation Rubric for RTI in Reading (I-RRR), to describe the rubric, which we have developed. This article focusing on addressing the following research questions:

Research Questions

1. To what extent does the I-RRR appropriately measure the construct of RTI implementation?
2. What evidence substantiates the reliability of the I-RRR for evaluating the implementation of each identified RTI component?

Methods

The development of I-RRR was adapted from standard procedures for developing test instruments, an approach selected with a goal to develop an instrument that could be used to assess RTI implementation. Six of Downing's (2006) recommended twelve-step process were used to develop the I-RRR as the other half of the steps specifically address tasks related to student assessments with individual items and scores. The six steps used to develop I-RRR include: (a) identifying the important components of the targeted domain; (b) operationalizing these important components and providing descriptors for each component at different levels of understanding; (c) creating test items based on the descriptors listed in the test blueprint; (d) gathering content related validity evidence; (e) conducting a pilot study on the assessment to gather evidence about the stability of the instrument, and (f) creating a technical report and documenting validity evidence (Downing, 2006).

The first step taken to construct the I-RRR was to identify the essential components of RTI through a synthesis of the literature. Second, operational definitions for each component were made to categorize implementation into three levels of implementation: *fully*, *partially* and *not at all*. Third, a set of descriptors, with observable or measurable criteria, was developed for each level of implementation within the I-RRR. Fourth, experts in the field of RTI reviewed and provided content validity evidence for the I-RRR. Fifth, two independent raters used the I-RRR to evaluate the implementation of RTI in two elementary schools. The inter-rater agreement between raters was established as well as evidence for the stability or reliability of the I-RRR. Finally, the raters' and the participating teachers' feedback on the relevance and ease of use of the I-RRR was gathered to document its utility and social validity. Combined, these steps provide information related to the validity and reliability of the I-RRR for use as a tool to evaluate RTI implementation.

For each component, an operational definition and three sets of descriptors are provided to describe the implementation of the components when being implemented fully, partially, or not being implemented at all. To be fully implemented all of the specified sub-components must be in place. A rating of 'partial implementation' indicates that most of the subcomponents were in place, and a rating of 'not implemented' indicated that none of the subcomponents were in place. By consulting the relevant literature, we selected the following criteria for each identified RTI component.

Content Experts Review the I-RRR

Continuing to follow the steps outlined by Downing and Haladyna (2006), once test items were developed, the test was sent to be reviewed three RTI experts. Drs. Ed Shapiro, Doug Marston, and Teri Wallace were invited to review the I-RRR because they had extensive experience in RTI implementation and had published numerous articles on RTI in peer-reviewed journals. In addition, all three experts had recently served as principal investigators for Model Demonstration Projects on RTI funded by the Office of Special Education Programs (OSEP).

These three RTI experts reviewed the I-RRR using a structured protocol. First, they reviewed the content to determine if there were any components of RTI not included in the I-RRR. Second, they indicated whether they agreed with the operational definitions and descriptors provided in the I-RRR. Finally, they provided reasons for disagreement and/or suggestions for revision.

Use the I-RRR to Evaluate RTI in Two Schools

After making slight revisions to the I-RRR based on expert reviewers' feedback and completed the detailed case study reports, two independent raters were recruited to use the I-RRR to evaluate RTI implementation at case study schools. The first independent rater was a project manager for a National Model Demonstration Center on RTI. She holds a Ph.D. in Educational Leadership from the University of Oregon. The second independent rater holds a doctoral degree from the Department of Special Education and Clinical Science at the University of Oregon. Selecting two raters of different backgrounds helped minimize the influence of possible shared bias.

After providing the independent raters with two copies of the revised I-RRR and school reports documenting information gathered at the case study schools, the raters were instructed on how to read and interpret the overall school report for the target students and asked to read the report independently and in its entirety before rating the quality of implementation. The first-author then reviewed the components and use of the I-RRR with them and answered any questions they had. For each component, raters indicated either 0 for *not implemented*, 1 for *partially implemented*, or 2 for *fully implemented* for each component.

Data Analysis

The I-RRR was developed in accordance to Kane's (2006) advice to integrate different strands of evidence to develop a plausible and coherent argument to represent the proposed interpretation and use of the scores in the test development process. Although the I-RRR is not a test *per se*, it is intended to be used to provide an assessment of the quality and extent of RTI implementation in elementary schools. Thus, this approach is warranted.

Measuring RTI Implementation

To answer our first research question (To what extent does the I-RRR appropriately measure the construct of RTI implementation?), we gathered experts' judgments as content related validity evidence for the validation of the I-RRR (Downing & Haladyna, 2006). First, the number of agreements and disagreements were calculated for each expert indicated on the I-RRR regarding the proposed operational definitions and descriptors. Second, the number of operational definitions and descriptors were calculated for which all three experts reached unanimous agreement. Third, the experts' feedback were compiled to make revisions to the I-RRR, bringing it in closer alignment to expert content reviewers' judgments about the essential components of RTI implementation. Revisions included clarifying the wording and deleting value-laden terms such as *appropriate* and *sufficient*. In addition, set time limits for instruction in Tier 2 and Tier 3 interventions were removed.

Reliability for Each RTI Component

To answer the second research question (What evidence substantiates the reliability of the I-RRR for evaluating the implementation of each identified RTI component?), we followed Miles and Huberman's (1994) recommendations for triangulating data from multiple sources. Sources included: interviews, observations, and document review. Following Yin's (2003) recommendation that researchers confirm their findings with key informants, member checking was conducted in two different ways: during the second interviews with participating teachers and administrators and by having them confirm the accuracy and completeness of the school reports compiled as part of the case study.

To provide evidence for the stability of the I-RRR, inter-rater agreement was conducted between the two independent raters who applied the I-RRR to the case study data presented in the detailed school reports. The two raters evaluated the level of implementation of each identified component using the revised I-RRR. They scored the school reports independently and had no interactions with each other. The first author analyzed Teacher 1's and Teacher 4's self-evaluation of implementation by reviewing their completed rubric and compared the teacher's ratings with the independent raters.

Preparation of School Reports

After the I-RRR had been reviewed and revised, case study was conducted to document the RTI implementation process in two elementary schools. Data from the case study were compiled into detailed school reports. The school reports were shared with teachers and administrators at both case study schools to verify accuracy and completeness prior to being shared with the independent raters, who applied the I-RRR to evaluate RTI implementation at the case study schools.

Results

Data analysis reveals following results.

Results for Question 1: Experts' Review of the I-RRR

The I-RRR was comprised of six identified components of RTI. Within each component, we provide an operational definition and three sets of descriptors, one for each level of implementation. Thus, each of the three RTI experts had 24 opportunities to state whether he or she agreed with the provided operational definition and descriptors. These three content experts made their judgments independently and did not meet to discuss or reconcile their differences.

Of the 24 provided definitions and descriptors, Dr. Ed Shapiro agreed with 18 (75%), Dr. Doug Marston agreed with 22 (92%), and Dr. Teri Wallace agreed with 23 (96%). In all, there were 63 agreements and 9 disagreements. The nine disagreements were scattered among 5 of the six components. The percentage of agreement across the components ranged from a low of 75% (9/12 on Tier 2 and Tier 3 interventions) to a high of 100% (12/12 on Progress Monitoring), with a mean agreement of 87.5%.

Revisions Made

Based on the experts' feedback, we made the following revisions:

First, we honored Dr. Shapiro's suggestion by deleting teacher-judgment from the operational definition of evidence-based decision-making. However, we did not explicitly state that teacher judgment should be banned from the evidence-based decision-making process. Dr. Shapiro took a conservative stand in interpreting evidence-based decision-making, and suggested that inclusion of teacher judgment would undermine the tenet of evidence-based decision-making. An equally plausible argument could be made for why teachers' professional judgment should be included in the evidence-based decision-making process. From this perspective, one might argue that just as doctors use their professional judgment to diagnose patients based on empirical evidence, teachers can use their professional judgment to diagnose students' skill deficits based on empirical evidence. Thus, explicitly prohibiting the inclusion of teacher judgment did not occur.

The value-laden terms from the I-RRR were either deleted or added an expected outcome as the standard of judgment. For example, operational definition of Tier 1 instruction was revised as,

'The comprehensive core reading instruction provided to all students in general education classes for a sufficient amount of time to meet grade-level instructional goals.' In Tier 1 Instruction, teacher teaches all "five big ideas" as specified in NRP or relevant ones to specific grade-level, using evidence-based teaching strategies.'

Whenever possible, the first author provided measurable criteria in the I-RRR to help users of the rubric make objective decisions. In the absence of quantifiable criteria, descriptions containing enough details to assist in differentiating full, partial and non-implementation of specific components were provided. For example, one of the descriptors for full implementation of organizational support was revised as follows:

Teachers receive multiple sessions of professional training and on-going support to implement research-based instruction, assessment and evidence-based decision-making. The professional training is focused and interactive. Teachers collaborate with colleagues to identify students' needs and implement tailored interventions to address the needs.

Its corresponding descriptor for partial implementation stated that

Teachers receive only initial professional training and have no on-going support to implement research-based instruction, assessment and evidence-based decision-making. Professional collaboration occurs in isolated incidences. The extent of collaboration is limited to issues of logistics and role responsibilities.

Its corresponding descriptor for non-implementation stated that

Teachers receive no professional training and rarely collaborate.

Revision following expert review is intended to be more than editing or rephrasing words; it is intended to ensure the validity of the content included on the I-RRR (Downing & Haladyna, 2006). It should be noted that the experts did not suggest additional components to be added to the I-RRR nor disagreed with the components provided, indicating support for the overall I-RRR. Revisions to the I-RRR followed expert reviewers' feedback prior to its use by the independent raters. The revised rubric can be found in Appendix A.

Results for Question 2: Reliability of the I-RRR

The I-RRR was used by two independent raters to evaluate two schools' RTI implementation. Each rater gave each school one rating score for the implementation of each identified RTI component. In all, they had 12 opportunities for agreeing or disagreeing with each other's appraisal. Both raters were in agreement that School 1 was fully implementing all of the components of RTI and that School 2 was fully implementing universal screening and partially implementing Tier 1 instruction, Tier 2 and Tier 3 interventions, evidence-based decision making and organizational support. The two raters agreed on 11 of the 12 occasions, only disagreeing on whether Progress Monitoring in School 2 was fully or partially implemented. The inter-rater agreement ($r = .92$) was calculated by dividing the number of agreements with the number of possible occasions.

Additional reliability evidence was gathered through the focus group, during which Teachers 1 and 4 conducted independent self-evaluations on the implementation of RTI at their grade-level using the I-RRR. Teacher 1 reported that she fully implemented RTI across all the six identified components, which is in agreement with the appraisal of both raters. Teacher 4 reported that the implementation of universal screening at her school was at the level of full implementation and the other five components were at the partially implemented level. She marked progress monitoring at the level of partial implementation because there was often a delay in receiving progress monitoring data from the school psychologist. Her appraisal is in 100% agreement with one rater and in 92% agreement with the other rater.

Discussion

The purpose of this study was to develop an assessment instrument to assist teachers and administrators to evaluate their implementation of RTI; however, developing a rubric to evaluate RTI is particularly challenging because neither research literature nor federal regulation specifies what constitutes RTI (Burns & Coolong-Chaffin, 2006), and the implementation of RTI is influenced by the individual nuances of the school and the district and community in which it is located (Berkeley et al., 2009; Coyne et al., 2001; Gradenet et al., 2007).

To overcome these two hurdles, we reviewed relevant RTI literature, identified six essential components of RTI, and developed an initial draft of the I-RRR based on these components. Three content experts reviewed the I-RRR, and revisions to the I-RRR were made based on their feedback. We conducted a field study to create school reports, describing how these components worked in practice in two different schools. Finally, two independent raters were selected to evaluate the implementation of RTI in two elementary schools, using the developed I-RRR and the school reports.

Major Findings of the Study

There are two major findings of this study. Each has significant implications for the validation of the I-RRR. First, the three experts provided content-related validity evidence by agreeing that the six identified components captured the essence of RTI and were sufficient to assess RTI implementation. Although there were some minor differences noted in terms of the operational definitions of the descriptors and how to qualitatively differentiate the level of implementation within each of the components, most of these differences revolved only around the operational definitions related to instruction, both Tier 1 as well as Tier 2 and 3. Second, the high

inter-rater agreement between the two raters and between the raters and participating teachers suggests that the description of RTI implementation in these two schools is credible and the I-RRR is stable.

Limitations of Current Project

As with any study, this project had limitations that impact generalization and interpretations of the findings. After the experts had provided feedback on the initial version of the rubric, modifications were made but the revised version of the I-RRR was not resubmitted to these same experts for a second review. As a result, we do not know if the revisions completely addressed the concerns of the experts or improved the overall quality of the I-RRR. Although the reviewers did not indicate the need for a second review of the I-RRR; an additional expert review might have provided a more complete validation process.

Prior to this study, there were no validated instruments to evaluate the implementation of RTI. In the absence of validated instruments, we followed Downing's advice and used the feedback of the content experts as a key source of content-related validity evidence. These three experts independently agreed with the provided operational definitions and descriptors for the six identified components, with some minor revisions. Their endorsement provides justification for including these "test items" in the I-RRR. The experts did not suggest the exclusion of any components or descriptors of the components. They did not suggest that any part of the I-RRR was irrelevant to an evaluation of the implementation of RTI. In addition, the experts did not indicate any part of the I-RRR that might bias against certain sub-groups of students. Most importantly, the experts did not identify any additional RTI components, nor did they suggest expanding the scope of the evaluation.

Conclusions

These findings suggest that the I-RRR encompasses all of the key components of RTI implementation. Messick (1989) defines "content integrity," where the items of a test assess all of the most important components of the targeted domain, as one of the key considerations for the validity of a test instrument. If an important component is excluded from the test, then the validity of the test is undermined because of construct under-representation (Webb, 2006). The fact that none of the content experts suggested any additional RTI components provides evidence to support the claim that the I-RRR satisfies the requirement of content integrity.

Based on evidence gathered from the literature review and content experts, the I-RRR matches the domain of RTI implementation at the elementary school level in the context of reading, and these six identified components are essential and sufficient to evaluate the implementation of RTI.

In this study, raters applied the I-RRR to evaluate the implementation of RTI at two different schools based on detailed school reports. The credibility of the school reports was substantiated by triangulating multiple data sources gathered during the field study and by confirming the findings with the stakeholders (participating teachers and administrators) during the second interviews (Miles & Huberman, 1994; Yin, 2003).

Part of evaluating reliability includes ensuring a standardized scoring procedure for an assessment instrument (Downing, 2006). In this study, the scoring procedure was written and provided to raters prior to their scoring. The two raters scored the I-RRR independently by reading the same school reports and scoring them one component at a time. Their notations on the I-RRR indicated that they both found justifiable cause to support their ratings. Their annotated rubrics provided evidence to support the claim that their ratings are fair and unbiased.

Implications for Practice

In this study, we provide foundational evidence for the Implementation Rubric of RTI in Reading (I-RRR) as an empirically sound criterion for successful implementation of RTI in elementary schools. This study serves as a base for which other research can continue to provide the empirical evidence necessary to strengthen the case for the I-RRR as a model for practitioners. Future research should substantiate the sensitivity for the I-RRR for measuring qualitative differences in schools' RTI implementation and whether or not these differences correspond with different student outcomes. The findings of this study, however, provide teachers with a technically adequate measure of six components of RTI implementation, identified in the literature and verified by experts in the field. Fully developing and validating the I-RRR as a model of best practice will require further attention, but this study provides an important first step in creating a valid and reliable way to measure a school-level policy that is currently being implemented to millions of children without any evidenced means of measuring the effectiveness of its implementation.

References

- Berkeley, S., Bender W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*, 85-95.
- Bollman, K. A., Silbergitt, B., & Gibbons, K. A. (2007). The St. Croix River education district model: Incorporating system-level organization and a multi-tiered problem solving Process for intervention delivery. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The handbook of Response to Intervention: The Science and Practice of Assessment and Intervention*, 319-330. New York: Springer.
- Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analysis review of responsiveness-to-intervention research: Examining field-based and research implemented models. *Journal of Psycho educational Assessments, 381-394*.
- Burns, M. K., & Coolong-Chaffin, M. (2006). Response to Intervention: The role of and effect on school psychology. *School Psychology Forum: Research in Practice, 1*(1), 3-15.
- Christ, T. J., & Hintze, J. M. (2007). Psychometric consideration when evaluating Response to Intervention. In S. R. Jimerson, M. K. Burns & A. Van Der Heyden (Eds.), *Handbook of Response to Intervention*, 93-105. New York, NY: Springer.
- 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), 204-206.
- Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2001). Prevention and intervention in beginning reading: two complex systems. *Learning Disabilities Research & Practice, 16*(2), 62-73.
- Darling-Hammond, L., & Mclaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*(8).
- Denton, C. A., Vaughn, S., & Fletcher, J. M. (2003). Bringing research-based practice in reading intervention to scale. *Learning Disabilities Research & Practice, 18*(3), 201-211.

- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 77, 433-480.
- Downing, S. T., & Haladyna, T. M. (Eds.). (2006). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum.
- Fletcher, J. M., Francis, D. J., Morris, R. D., & Lyon, G. R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 34(3), 506.
- Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all Children. *Learning Disabilities Research & Practice*, 16(4), 203-212.
- Fuchs, L. S. (1995). *Curriculum-Based Measurement and Eligibility Decision-Making: An Emphasis on Treatment Validity and Growth*. Washington, DC: National Research Council.
- Fuchs, L. S. (1998). Treatment fidelity: A unifying concept for reconceptualizing the identification process of learning disabilities. *Journal of Learning Disabilities*, 13, 2014-219.
- Fuchs, L. S. (2003). Assessing intervention responsiveness: conceptual and technical issues. *Learning Disabilities Research & Practice*, 18(3), 172-186.
- Fuchs, D., Compton, D.L., Fuchs, L.S., Bryant, J., & Davis, N.G. (2008) Making “secondary intervention” work in a here tier responsiveness-to-intervention model: Findings from the first grade longitudinal readings study of the National Research Center on Learning Disabilities. *Reading and Writing Quarterly: An interdisciplinary Journal*, 21 (4), 413-436.
- 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(2), 129-136.
- Fuchs, D., Fuchs, L.S., & Compton, D.L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, 78(3), 263-269.

- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-Intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18*(3), 157-171.
- Gersten, R., Beckman, S., Clarke, B., Foegen, A., Marsh, L., Star, J., Scott, L. (2009). *Assisting students struggling with mathematics: Response to intervention (RTI) for elementary and middle schools (Institute of Education Sciences Practice Guide)*. Washington, D.C.: U.S. Department of Education.
- 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. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV*, 679-700.
- Good, R. H., Simmons, D., Kame'enui, E., & Chard, D. (2003). *Operationalizing Response to Intervention in eligibility decisions*. Paper presented at the National Research Center on Learning Disabilities Response to Intervention Symposium, Kansas City, MO.
- Graden, J. L., Stoller, S. A., & Poth, R. L. (2007). The Ohio Integrity System Model: an overview and lessons learned. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention*, 288-299. New York: Springer.
- Greenwood, C. R., Abott, M., & Tapia, Y. (2003). Ecobehavioral strategies: observing, measuring, and analyzing behaviors and reading interventions. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom*. Baltimore, NY: Brooks.
- Gresham, F. M., Reschly, D. J., Tilly, W. D., Fletcher, J., Burns, M., & Prasse, D. (2005). A response to intervention perspective. *The School Psychologist, 59*(1), 26-33.
- Grimes, J., & Kurns, S. (2003). *An intervention-based system for addressing NCLB and IDEA expectations: A multiple tiered model to ensure every child learns*. Paper presented at the National Research Center on Learning Disabilities Responsiveness to Intervention Symposium Kansas, MO.

- Harn, B., Kame'enui, E. J., & Simmons, D. C. (2007). The role of the third tier. In D. Hager, J. Klinger & S. Vaughn (Eds.), *Evidence-based reading instruction for Response to Intervention* (pp. 161-184). Baltimore: Paul H. Brooks.
- Hoover, J.J., Baca, L, Wexler-Love, E., & Saenz, L.S., (2008). *National Implementation of Response to Intervention (RTI): Research Summary*. BUENO Center-School of Education, University of Colorado. Retrieved from <https://www.nasdse.org/Portals/0/NationalImplementationofRTI-ResearchSummary.pdf>
- Hughes, C. & Dexter, D.D. (2016). *Field Studies of RTI Programs, Revised*. Retrieved from: <http://www.rtinetwork.org/learn/research/field-studies-rti-programs>
- Hurlbut, A.R. & Tunks, J. (2016). Elementary pre-service teachers' experiences with Response to Intervention. *Teacher Education Quarterly*, 43(3), 25-48.
- Jenkins, J. R. (2003). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities (NRCLD): Response to Treatment as Learning Disabilities Identification, Kansas, MO.
- Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). *Responsiveness to intervention: How to do it*. Lawrence, KS: National Research Center on Learning Disability.
- Justice, L. M. (2006). Evidence-based practice, Response to intervention, and the prevention of reading difficulties. *Language Speech Hearing Serving School*, 37(4), 284-297.
- Kamps, D. M., & Greenwood, C. R. (2003). *Formulating secondary level reading interventions*. Paper presented at the National Research Center on Learning Disabilities Response to Intervention Symposium, Kansas, MO.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education / Praeger.
- Lau, M. Y., Sieler, J. D., Muyskens, P., Canter, A., VanKeuren, B., & Marston, D. (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large midwest urban school district. *Psychology in the Schools*, 43(1), 117-127.

- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*, 129-151.
- Marston, D. (2005). Tiers of intervention in Responsiveness to Intervention: prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities, 38*(6), 539.
- Marston, D., Lau, M., & Muyskens, P. (2007a). Implementation of Problem-Solving Model in the Minneapolis Public Schools. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The Handbook of response to intervention: The science and practice of assessment and intervention*, 279-287. New York: Springer.
- Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-Solving Model for decision making with high-incidence disabilities: the Minneapolis experience. *Learning Disabilities Research & Practice, 18*(3), 187-200.
- Marston, D., Pickart, M., Reschly, A., Heistad, D., Muyskens, P., & Tindal, G. (2007b). Early literacy measures for improving student reading achievement: translating research into practice. *Exceptionality, 15*(2), 97.
- McMaster, K. L., & Wagner, D. (2007). Monitoring response to general education instruction. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The Handbook of Response to Intervention: The science and practice of assessment and intervention*, 223-233. New York: Springer.
- Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research and Practice, 25*, 217-225.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.), 13-103. New York: American Council on Education, MacMillan.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded source book*. (2nd ed.). Thousand Oaks, CA: Sage.
- National Reading Panel. (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH pub. 00-4754): National Institute of Health.

- 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*, 220-235.
- O'Connor, R., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: change in teaching and student outcomes. *Journal of Learning Disabilities, 38*, 440-445.
- O'Connor, R., Harty, K. R., & Fulmer, D. (2005). Tiers of interventions in kindergarten through third grade. *Journal of Learning Disabilities, 38*, 532-538.
- Peterson, D. W., Prasse, D. P., & Shinn, M. R. (2007). The Illinois flexible service delivery model: a problem-solving model initiative. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The handbook of Response to Intervention: the science and practice of assessment and intervention*, 300-318. New York: Springer.
- Reschly, A., Coolong-Chaffin, M., Christenson, S. L., & Gutkin, T. (2007). Contextual influences and Response to Intervention: critical issues and strategies. In S. R. Jimerson, M. Burns & A. VanDerHeyden (Eds.), *The handbook of Response to Intervention*. New York: Springer.
- Shapiro, E. S., Zigmond, N., Wallace, T., & Marston, D. (Eds.). (2011). *Models for implementing response to intervention: Tools, outcomes, and implications*. New York, NY: Guilford.
- Simmons, D. C., Coyne, M.D., Hagan-Burke, S., Kwok, O., Johnson, C., Zuo, Y., Crevecoeur, Y.C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children, 77*, 207-228.
- Spectrum K12. (2010). *Response to Intervention adaption survey 2010*. Spectrum K12. Retrieved from rti.pearsoned.com/docs/RTISite/2010RTIAdoptionSurveyReport.pdf
- Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18*(3), 147-156.

- Spencer, T.D., Petersen, D.B., & Adams, J.L. (2015). Tier 2 language intervention for diverse preschoolers: An early-stage randomized control group study following an analysis of response to intervention. *American Journal of Speech-Language Pathology, 24*, 619-638.
- Stahl, K.A.D.; Keane, A.E., Simic, O. (2013). Translating policy to practice: Initiating RTI in urban schools. *Urban Education, 48*, 350-379.
- Tindal, G., Yovanoff, P., & Alonzo, J. (2006). *DIBELS in the context of reform and the measurement of reading* (Vol. 1). Newport, OR: The Advocate, a publication of Oregon Reading Association.
- Torgesen, J., Alexander, A., Wagner, D., Rashotte, C. A., Voeller, K., & Conway, T. (2001). Intensive remedial reading instruction for children with severe reading disabilities: immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33-58.
- Van Der Heyden, A. M. (2011). Technical adequacy of Response-to-Intervention decisions. *Council for Exceptional Children, 77*, 335-350.
- Van Der Heyden, A. M., & Jimerson, S. R. (2005). Using Response-to-Intervention to enhance outcomes for children. *The California School Psychologist, 10*, 21-32.
- Van Der Heyden, A., Witt, J. C., & Naguin, C. (2003). The development and validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204-207.
- Vaughn, S., Cirino, P.T., Wanzek, J. Wexler, J., Fletcher, J.M., Denton, C.A., Francis, D.J. (2010). Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. *School Psychology Review, 39*, 3-21.
- Vaughn, S., & Denton, C. A. (2008). Tier 2: The role of intervention. In D. Fuchs, L. S. Fuchs, & S. Vaughn (Eds.), *Response to Intervention: A Framework for Reading Educators* (pp. 51-70). Newark, DE: International Reading Association.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: To promise and potential problems. *Learning Disabilities Research & Practice, 18*(3), 137-146.

- Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to Instruction as a means of identifying students with reading/learning Disabilities. *Exceptional Children, 69*(4), 391-409.
- Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C., Barth, A., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities, 45*, 515-525.
- Webb, N. R. (2006). Content for achievement tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 155-180). Mahwah, N. J.: Lawrence Erlbaum.
- Yin, R. K. (2003). *Case study research: Design and methods* (3 ed.). Thousand Oaks, CA.