



What's Happening

November 2016

Measuring the implementation fidelity of the Response to Intervention framework in Milwaukee Public Schools

Stephen J. Ruffini
Ryan Miskell
WestEd

Jim Lindsay
Maurice McInerney
Winsome Waite
American Institutes for Research

In collaboration with the Midwest Urban Research Alliance

Key findings

Regional Educational Laboratory Midwest assisted Milwaukee Public Schools in developing a system for measuring schools' progress in implementing Response to Intervention (RTI), a pedagogic method that uses tiered levels of instruction adapted to student needs. This study examined the ratings produced by that system in 2014/15 to determine the system's reliability, schools' progress in implementing RTI, and any relationship of the ratings to school characteristics. The study found the following:

- The district-customized rubric for measuring the implementation fidelity of the RTI framework showed good interrater and interitem reliability.
- Some 53 percent of participating elementary schools that were rated using the rubric were implementing RTI with adequate fidelity. Schools with the lowest academic performance (priority schools) struggled most with implementation.
- Among components of the RTI framework, schools struggled most with multitiered instruction and evaluation.
- Implementation fidelity ratings were related to the percentage of teachers with advanced credentials, retention rates of licensed staff, percentage of economically disadvantaged students, and percentage of students suspended during the school year.



Institute of Education Sciences
U.S. Department of Education



U.S. Department of Education

John B. King, Jr., *Secretary*

Institute of Education Sciences

Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance

Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
Elizabeth Eisner, *Project Officer*

REL 2017–192

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

November 2016

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0004 by Regional Educational Laboratory Midwest administered by the American Institutes for Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Ruffini, S. J., Lindsay, J., McInerney, M., Waite, W., & Miskell, R. (2016). *Measuring the implementation fidelity of the Response to Intervention framework in Milwaukee Public Schools* (REL 2017–192). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

This report is available on the Regional Educational Laboratory website at <http://ies.ed.gov/ncee/edlabs>.

Summary

Many schools identified by states as needing improvement through their Elementary and Secondary Education Act waivers have selected Response to Intervention (RTI), a three-tiered instruction program sometimes referred to as tiered levels of instruction, as one of their main strategies for improving school performance and closing achievement gaps. Yet research on the effects of tiered interventions in school settings is thin (Gersten et al., 2008). Most studies that show strong impacts have focused on small samples of schools where the leaders of small group instruction (tier 2 instructors) were employed by the intervention developers, thereby allowing the developers to pay close attention to the quality of implementation and to give direct guidance on tier 2 instruction (for example, Fuchs et al., 2005). Studies of these same interventions that involve more schools and that use school staff to lead the small group instruction often find smaller effects. Several factors may explain why the larger studies produce smaller effects; one such factor may be that little effort was made by the schools to monitor implementation systematically and use implementation information as the basis for improvements (Rolfhus et al, 2012).

Prior to this study Regional Educational Laboratory Midwest and partners affiliated with the former National Center on Response to Intervention worked with Milwaukee Public Schools to develop a research-based rubric for rating school-level implementation of RTI. The rubric was coupled with a data dashboard that analyzes ratings, displays results at various levels of aggregation, and identifies RTI components that are being implemented inadequately and require improvement. National Center on Response to Intervention staff successfully trained 22 of the district's school improvement coaches to use the rubric when rating implementation during school visits and to enter the ratings into the data dashboard.

The current study analyzed ratings by district staff employed as school improvement coaches and who volunteered for the study. In 2014/15 these school improvement coaches visited 70 district schools that serve students in grades K–5. The coaches examined documents and interviewed school staff on implementation of RTI. Based on the information they gathered during a school visit, the coaches rated the schools' implementation of RTI using a 33-item rubric. Ratings for two schools were incomplete, leaving 68 schools in the sample. Analyses focused on the reliability of the RTI implementation rubric, average implementation ratings across the 68 schools, and correlations between aggregate ratings and school characteristics.

Key findings include the following:

- Ratings of the same schools made independently by two school improvement coaches employed by the district showed a high degree of consistency, even after accounting for chance (Cohen's kappa interrater reliability estimates range from .71 to .85 for the various components).
- Ratings across the 33 indicators in the implementation fidelity rubric showed a high degree of consistency ($\alpha = .94$), and the consistency of ratings on indicators for the six key RTI components fell in the adequate or good range (alphas for key components ranging from .70 to .85).
- Two years after rolling out RTI, all 68 schools had made progress toward implementing the framework, and 53 percent of schools were found to be implementing it with fidelity.

- Some 69 percent of schools had yet to implement the multitiered instruction component with adequate fidelity (especially the tier 3 subcomponent), and 49 percent had yet to implement the evaluation component with adequate fidelity. These components were subsequently identified as priority areas for additional school-level professional development and coaching within the district.
- Several school characteristics showed statistically significant relationships with implementation ratings for RTI components. Specifically, higher-performing schools and schools with higher percentages of teachers with advanced credentials, higher staff retention rates, smaller percentages of economically disadvantaged students, and lower student suspension rates showed stronger implementation of RTI than did schools without these characteristics.

Contents

Summary	i
Why this study?	1
Monitoring Response to Intervention	1
Milwaukee Public Schools' system for monitoring the implementation fidelity of Response to Intervention	3
Broader applications of this work	4
What the study examined	5
What the study found	8
Ratings on implementation of the Response to Intervention framework were reliable	8
Some 53 percent of schools visited showed adequate implementation fidelity for the Response to Intervention framework	10
Among the subcomponents of multitiered instruction, most schools had not implemented tier 3 prevention or appropriate instruction for culturally and linguistically diverse students with fidelity	10
Implementation ratings for the Response to Intervention framework were related to characteristics of school teacher and student populations	13
Implications of the study findings	15
Limitations of the study	16
Appendix A. The district's implementation fidelity monitoring system for the Response to Intervention framework	A-1
Appendix B. Data collection and methodology	B-1
Notes	Notes-1
References	Ref-1
Boxes	
1 What is Response to Intervention?	1
2 Key terms	4
3 Milwaukee Public Schools' implementation fidelity monitoring system for the Response to Intervention framework	5
4 Data and methods	7
Figures	
1 Interrater reliability was good for the entire implementation fidelity monitoring rubric for the Response to Intervention framework and for each component for the sample of Milwaukee Public Schools, 2014/15	9
2 Interitem reliability met or exceeded the benchmark for adequate reliability for the entire implementation fidelity monitoring rubric for the Response to Intervention framework and for each component for the sample of Milwaukee Public Schools, 2014/15	9

3	More than half the Milwaukee Public Schools sample had adequate implementation fidelity overall for the Response to Intervention framework, 2014/15	11
4	Adequacy of implementation fidelity for the Response to Intervention framework in the Milwaukee Public Schools sample was lowest among priority schools, followed by focus schools and other schools, 2014/15	12
5	Most schools in the Milwaukee Public Schools sample had not implemented tier 3 instruction or appropriate instruction for culturally and linguistically diverse students with fidelity, 2014/15	13

Tables

1	Average implementation fidelity ratings for the Response to Intervention framework for the Milwaukee Public Schools sample, by school type, 2014/15	11
2	Correlations between implementation fidelity ratings for the Response to Intervention framework and contextual factors for Milwaukee Public Schools, 2014/15 (Pearson product-moment correlation coefficients)	14
A1	Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5	A-3
B1	Contextual factors that might be related to the implementation fidelity of the Response to Intervention framework	B-2

Why this study?

Response to Intervention (RTI), a three-tiered program of instructional support, has been widely adopted as a framework for meeting the instructional needs of students and as a school improvement strategy (see box 1). In a 2011 survey 68 percent of U.S. public school districts indicated that they had implemented or were in the process of implementing RTI as a strategy to improve learning and achievement among all students, including high-need students struggling with basic math and reading skills (Detgen, Yamashita, Davis, & Wraight, 2011; Global Scholar et al., 2011 [as cited by Shah, 2011]; National Center on Response to Intervention, 2010). In addition, 40 of 43 states or jurisdictions with approved Elementary and Secondary Education Act flexibility waivers (as of September 2015, including the District of Columbia and Puerto Rico) explicitly mention tiered levels of instructional support for students as a primary approach to improving low-performing schools.

Monitoring Response to Intervention

RTI is a data-based decisionmaking approach to instruction in which teachers determine the amount of instruction (tiers of support) that students need in a subject (typically math or reading) based on their performance on a screening assessment (Fuchs, Fuchs, & Compton, 2012). Tier 1 involves core classroom instruction for all students. Students who perform poorly on a subject-matter screening assessment become eligible for tier 2 supplemental instruction or intervention in that subject (Deno, 1985; Vaughn, Denton, &

RTI is a data-based decisionmaking approach to instruction in which teachers determine the amount of instruction that students need in a subject (typically math or reading) based on their performance on a screening assessment

Box 1. What is Response to Intervention?

Response to Intervention (RTI) integrates assessment and intervention within a multilevel prevention framework to maximize student achievement and reduce behavioral problems. Schools identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions, and adjust the intensity and nature of those interventions on the basis of a student's responsiveness (National Center on Response to Intervention, 2010, pp. 1–2).

RTI's distinctive feature is its data-based decisionmaking approach (Fuchs et al., 2012) combined with tiers of support for students, depending on their needs. Tier 1 involves core classroom instruction for all students in the focus subjects (such as math and reading). Students who test poorly in those subjects then become eligible for tier 2 supplemental small group instruction in the subjects (Deno, 1985; Vaughn et al., 2010). Tier 3 provides more individualized and intensive instruction for students who do not respond to tiers 1 and 2.

Researchers affiliated with National Center on Response to Intervention¹ identified the following five component processes of successful implementation of RTI:

- Screening.
- Multilevel prevention/intervention.
- Progress monitoring.
- Data-based decisionmaking.
- Overarching factors (focus on prevention, leadership, staff qualifications, cultural and linguistic responsiveness, communication with parents).

Note

1. The National Center on Response to Intervention was a technical assistance center that supported states and school districts' efforts at establishing RTI. It ran from 2007 to 2012 through a grant from the U.S. Department of Education's Office of Special Education Programs.

Fletcher, 2010). Tier 3 involves more individualized and intensive instruction for students who do not respond to tiers 1 and 2.

The Institute of Education Sciences of the U.S. Department of Education has published practice guides on using RTI to improve student achievement in math (Gersten et al., 2009) and reading (Gersten et al., 2008). Both practice guides report strong evidence that tier 2 interventions that are systematic, explicit, and focused on students' skill deficits improve the academic achievement of students whose performance is below expectations in the early elementary grades. The recommendations in the practice guides have strong empirical research support, but many of the studies were conducted under particularly favorable conditions, such as having the interventionists who led the small group supplemental instruction be in the employ of the program developer and allowing the program developer to monitor implementation fidelity and provide interventionists with strategies for improving the quality of implementation. Little evidence is available to determine whether the strong rating reported in the practice guides would persist if RTI's tiered instruction approach were implemented within real school contexts with school staff serving as interventionists.¹

Assessing the quality of implementation of RTI matters when attempting to examine the impacts of interventions in school settings with school or district staff providing the interventions

Two recent rigorous studies that examined the same tier 2 intervention produced different results, which may be explained by differences in the settings in which the intervention was implemented and differences in the intensity of fidelity monitoring in the studies. Both studies examined the impact of Number Rockets[®], a tier 2 intervention for math in grade 1 (Fuchs et al., 2005; Rolfhus et al., 2012). In one study the tier 2 intervention was carried out by interventionists employed by the program developer, and the program developers were able to closely monitor the implementation fidelity and provide continuous feedback to the interventionists based on that monitoring (Fuchs et al., 2005). The study found statistically significant positive effects on four tests measuring computation and math concepts but no statistically significant differences on tests of applied problems and fact fluency. In the second study, which included a larger number of schools, school staff served as interventionists and received little feedback on implementation fidelity during the study (Rolfhus et al., 2012). That study also reported statistically significant positive results on the same four tests measuring math skills, but the effects were smaller than in the 2005 study. The schools in the 2012 study also showed less implementation fidelity than did those in the 2005 study (Fuchs et al., 2005). The 2012 report noted that the lower fidelity may have contributed to the smaller effect found in that study compared with that observed in the 2005 study.

The importance of examining implementation fidelity when monitoring RTI has also been highlighted in efforts to explain null effects across numerous large-scale randomized controlled trials (for example, Hulleman & Cordray, 2009). The message that has emerged is that assessing the quality of implementation matters when attempting to examine the impacts of interventions in school settings with school or district staff providing the interventions. Any state or district that seeks to generate impacts similar to those from small-scale, tightly controlled studies could benefit from using a fidelity monitoring system to identify the parts of interventions that need stronger implementation.

The complexity of the RTI framework and the number of processes involved highlight the need to monitor the implementation fidelity of all component processes (Keller-Margulis, 2012). These component processes include the assessment process for student performance,

the instruction itself, and the decisionmaking processes (Keller-Margulis, 2012). Proponents of RTI refer to component processes as mechanisms: tracking daily instruction, providing reading coach support, and providing models of instructional implementation (Bianco, 2010). The National Center on Response to Intervention has identified five component processes (see box 1).

State education agencies and federally funded technical assistance centers have developed self-assessment measures to help schools and districts monitor the implementation fidelity of RTI. For example, the Florida Department of Education developed a self-assessment to measure implementation fidelity on six domains. The Self-Assessment of Multi-Tiered System of Supports Implementation focuses on measuring the critical components of multi-tiered systems of support so they can be implemented and sustained with fidelity (Florida Department of Education, 2014). Similarly, the Wisconsin RTI Center (2015) developed the School-wide Implementation Review, which is a self-assessment that school leaders complete to measure implementation of the four components of the state-recommended RTI framework.² The National Center on Response to Intervention also developed the RTI Framework Integrity Rubric and the RTI Framework Integrity Worksheet that can be used by outside raters who complete the worksheet during school visits or by school administrators as a self-assessment (National Center on Response to Intervention, 2010).

Milwaukee Public Schools' system for monitoring the implementation fidelity of Response to Intervention

Milwaukee Public Schools has included RTI within its Corrective Action Plan and has been rolling out the tiers for reading and math since 2012/13. Through its connection with the Midwest Urban Research Alliance, Milwaukee Public Schools requested a way to obtain objective information on whether its schools were implementing the RTI framework correctly and whether parts of the RTI framework needed to be improved.³ To address this need, Regional Educational Laboratory Midwest partnered with the district and other members of the Midwest Urban Research Alliance to develop a system with three features: the capability to provide the district and schools with unbiased formative information about how to improve implementation of RTI, the capability to provide the implementation information needed by the Wisconsin Department of Public Instruction, and appeal to school and district staff (that is, it must be viewed as a useful tool so that school and district staff charged with implementing RTI are likely to use it).

The system consists of three parts:

- A customized 33-indicator rubric that assesses the degree to which component processes are being implemented with fidelity to the RTI framework (see appendix A for the rubric).
- The process and materials for training school improvement coaches who were employed by the district and tasked with visiting schools and generating implementation ratings based on the evidence collected.
- A data dashboard that collects school improvement coaches' ratings, aggregates the ratings for all indicators in the rubric as well as those for each component process, displays the aggregated scores, and highlights the components that need additional work.

Between November 2014 and June 2015 school improvement coaches employed by Milwaukee Public Schools visited 70 schools and generated ratings using the rubric and evidence

Regional Educational Laboratory Midwest partnered with Milwaukee Public Schools and other members of the Midwest Urban Research Alliance to develop a system to provide objective information on whether schools were implementing the district's RTI framework correctly and whether parts of the RTI framework needed to be improved

obtained at the schools. This study examined the data that emerged from the school improvement coaches' use of this monitoring system in 68 schools (ratings for two schools were incomplete). The data were analyzed to determine whether the school improvement coaches had been properly trained to understand RTI components and use the implementation monitoring system reliably. In addition to discussing the results of those analyses, this report summarizes how well schools in the Milwaukee Public Schools district are implementing RTI generally and identifies which RTI components should be emphasized in professional development or coaching sessions. Finally, the study team calculated correlations between school characteristics and implementation ratings to determine whether school characteristics were related to stronger implementation.

Broader applications of this work

While this report presents findings that are specific to Milwaukee Public Schools, the description of the development of the implementation fidelity monitoring system for the RTI framework (see appendix A) can help administrators in other districts and states create their own monitoring system. The reliability results (see box 2 for definitions of key terms) can inform them about the consistency of the results obtainable when the implementation fidelity rubric is used by trained district staff. Moreover, the findings indicate that such a system can show not just how well schools are implementing RTI generally but also which of its components and subcomponents should be the focus of professional development. The correlation analysis can also inform school administrators about factors that might contribute to implementation fidelity. However, correlation findings are suggestive at best and cannot be used to infer causal relationships.

The description of the development of the implementation fidelity monitoring system for the RTI framework in this report can help administrators in other districts and states create their own monitoring system

Box 2. Key terms

Fidelity. A classification signifying the degree to which an intervention or program is implemented as intended. To determine the implementation fidelity of the Response to Intervention (RTI) framework, the study team adopted the implementation score cutpoints recommended by partners affiliated with the National Center on Response to Intervention. The cutpoints define the following fidelity categories:

- *Little fidelity* (average ratings less than 2.00). Schools that have made little progress in implementing RTI.
- *Inadequate fidelity* (average ratings of 2.00–3.49). Schools that have made some progress in implementing RTI but whose progress is inadequate for schools that have been implementing RTI for two years.
- *Adequate fidelity* (average ratings of 3.50–4.99). Schools that have implemented RTI at a level considered reasonable for two years of implementation.
- *Full fidelity* (average ratings equal to 5.00). Schools that were given the highest possible ratings for indicators making up components.

Implementation rating. A numeric value ranging from 1 to 5 assigned by a school improvement coach for each indicator based on the degree to which the indicator was present.

Reliability. The consistency of ratings. For this study examining school improvement coaches' ratings for indicators in the rubric, two types of reliability were examined:

- *Interrater reliability* is the consistency of ratings given to the same schools by different raters (the amount of measurement error due to differences among raters).

(continued)

Box 2. Key terms (continued)

- *Interitem reliability* is the consistency of ratings for indicators that presumably reflect the same construct (the amount of measurement error due to differences among the indicators).

Rubric. A tool that can be used by observers to determine the degree to which certain criteria are met. See box 3 and appendix A for descriptions of the rubric used in this study.

School types. The study looked at three types of schools: priority, focus, and other. *Priority schools* and *focus schools* are labels given to schools by the Wisconsin Department of Public Instruction under the state's Elementary and Secondary Education Act waiver.

- *Priority schools* are Title I schools in which overall student achievement was in the lowest 5 percent of Title I schools in the state.
 - *Focus schools* are Title I schools in which overall student achievement was in the lowest 10 percent of Title I schools in the state and either subgroup performance was very low or achievement gaps between subgroups were the most significant.
 - *Other schools* either were not Title I schools or were not classified as priority or focus schools.
-

What the study examined

This study examined the ratings by school improvement coaches employed by Milwaukee Public Schools who were trained to rate schools' implementation of RTI using the implementation fidelity monitoring system (box 3). The study team calculated reliability statistics for the ratings, average ratings for each component of RTI, and average ratings across the entire implementation fidelity rubric. If those preliminary statistics suggested that the rubric was reliable, the study team then used schools' average ratings to classify their level of implementation as showing little fidelity, inadequate fidelity, adequate fidelity, and full fidelity (see box 2).

Box 3. Milwaukee Public Schools' implementation fidelity monitoring system for the Response to Intervention framework

Overview of the rubric for assessing implementation fidelity

The district's implementation fidelity monitoring system for the Response to Intervention (RTI) framework was created by Regional Educational Laboratory Midwest, partners affiliated with the former National Center on Response to Intervention, and Milwaukee Public Schools and revised on the basis of feedback from district staff (see appendix A). The system includes a 33-indicator rubric. Ratings on the rubric's indicators can be aggregated into a single average rating representing the quality of implementation of the RTI framework as a whole and into average ratings for each of the following six components:

- Data-based decisionmaking (2 indicators).
- Balanced assessment¹ (6 indicators grouped into three subcomponents).
 - Screening (3 indicators).
 - Progress monitoring (2 indicators).
 - Culturally and linguistically responsive assessment (1 indicator).
- Multitiered instruction (14 indicators grouped into four subcomponents).
 - Tier 1 core curriculum (4 indicators).
 - Tier 2 prevention (5 indicators).

(continued)

Box 3. Milwaukee Public Schools' implementation fidelity monitoring system for the Response to Intervention framework *(continued)*

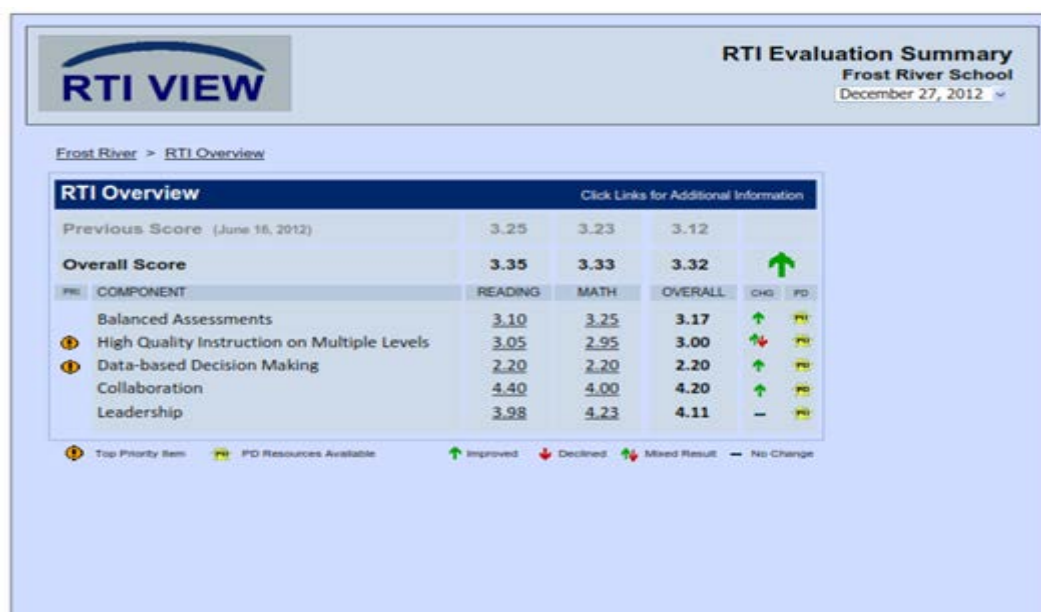
- Tier 3 prevention (4 indicators).
- Culturally and linguistically responsive instruction (1 indicator).
- Leadership (6 indicators).
- Collaboration (3 indicators).
- Evaluation (2 indicators).

The district's school improvement coaches were trained to base their implementation ratings on the types of evidence obtained during their visit to the school. The rubric included descriptors of the type of evidence that must be present to attain ratings of 1, 3, and 5 (with higher ratings corresponding to greater presence of the indicator). Ratings of 2 and 4 could also be assigned if evidence suggested a level of implementation between two other ratings. As an example, for the indicator titled "data system" under the data-based decisionmaking component, the descriptor for a rating of 1 was "no data system is in place to document and access individual student-level data"; the descriptor for a rating of 3 was "a data system is partially in place to document and access individual student-level data"; and the descriptor for a rating of 5 was "a comprehensive data system is in place to document and access individual student-level data." However, if a school's data system contained students' benchmark data (but no screening data), and only a school administrator had access to that data, then the school improvement coach might assign a rating of 2 for that indicator for that school.

The data dashboard

The implementation fidelity monitoring system also included a data dashboard for storing ratings and aggregating them across indicators, subcomponents, and components for each school. The dashboard's data display interface allowed administrators in schools, regions within the district, and the district to view average ratings for components, subcomponents, and indicators (see sample data display below). This display also identified priority items (components that need additional professional development of school and district staff).

Sample data display



Note

1. Refers to use of assessments for summative purposes (to determine students' level of understanding and skill development) and formative purposes (to determine whether students are eligible for tier 2 or tier 3 instruction and to monitor their progress).

Three research questions guided the study:

- How reliable is the district's implementation monitoring system for the RTI framework?
- To what extent are participating elementary schools implementing the RTI framework with fidelity?
- Are schools' implementation scores statistically related to school characteristics, such as teacher characteristics, student characteristics, and other schoolwide factors?

To address the research questions, the study team assessed the reliability of the rubric (the amount of consistency in ratings between the two school improvement coaches who visited the same schools and the amount of consistency in ratings for indicators representing the same constructs). Next, the study team looked at the descriptive statistics produced by the system to gauge average implementation ratings and fidelity classifications districtwide, within specific types of schools, and within specific component processes. The team also examined the relationships between aggregated ratings for the components and school characteristics (analytic methods are described in box 4 and appendix B).

Box 4. Data and methods

School visits

Milwaukee Public Schools officials recruited 70 schools serving students in grades K–5 to volunteer for the study. This is the maximum number of schools that district resources could support in a school year. To participate, school principals had to consent to allow school improvement coaches to visit their schools and gather information about their schools' implementation of RTI.

The district identified 23 school improvement coaches to visit schools and rate implementation using the rubric. They received three days of training from former consultants with the National Center on Response to Intervention who worked with Milwaukee Public Schools to develop the fidelity monitoring system. All but one school improvement coach who participated in the training passed the certification test. Each school improvement coach completed from 1 to 31 school visits and sets of ratings. Two school improvement coaches conducted each school visit. These visits were conducted from November 2014 to June 2015, with school visits by the two coaches occurring within a week of each other. During the visits the school improvement coaches examined documents, interviewed staff, observed tier 2 instruction, rated the schools on each of the rubric's 33 indicators, and entered the ratings into the data dashboard. The two school improvement coaches for each school compared their ratings and reconciled discrepant ratings. The individual ratings and reconciled ratings were the data analyzed for the study.

Contextual factors

Characteristics of schools in the sample were downloaded from the Wisconsin Department of Public Instruction website that included data on teacher and student characteristics and other schoolwide factors. District staff provided additional data for teacher characteristics.

Analytic methods

The study team conducted an exploratory analysis of the data, including checking the completeness of the ratings, the number of school improvement coaches that visited each school, and the subjects reviewed. Of the 70 schools visited, ratings were completed for 68 schools.

(continued)

Box 4. Data and methods *(continued)*

To address the first research question on reliability, interrater reliability was estimated using the number of consistent ratings divided by the total number of ratings. Cohen's kappa statistics were also calculated to account for chance ratings. Estimates of interitem reliability were calculated using Cronbach's alpha.

The second research question on how well schools are implementing RTI was addressed by calculating average ratings for each school on the overall rubric and on each of the six components. These average ratings were used to classify schools' implementation as showing little fidelity, inadequate fidelity, adequate fidelity, or full fidelity.

The third research question on the relationships between the overall level of implementation and factors related to teachers, students, and schoolwide achievement was addressed by calculating Pearson product-moment correlation coefficients. The Benjamini–Hochberg correction was also conducted to account for the likelihood of false positive results (Benjamini & Hochberg, 1995; for more details on methods, see appendix B).

What the study found

The first set of findings in this section addresses whether the rubric and monitoring system produced reliable data. The second set of findings are the implementation scores and classifications of implementation fidelity for the 68 schools overall and for the three types of schools (see box 2). Finally, to gain insight into characteristics that may be related to a school's ability to implement RTI, the correlations between school characteristics and RTI implementation ratings are presented.

Ratings on implementation of the Response to Intervention framework were reliable

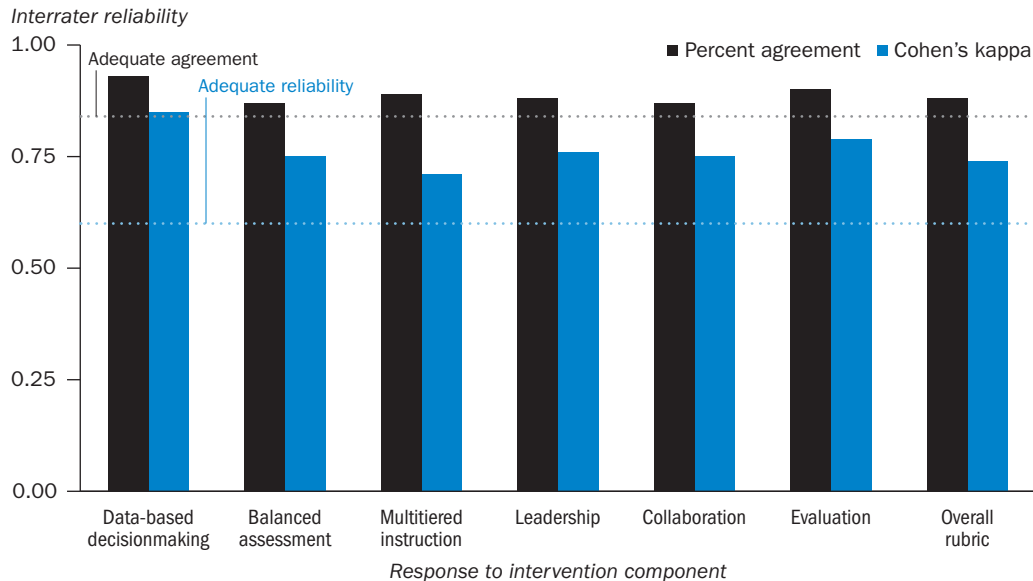
For this study reliability represents the consistency of ratings or the degree to which error from various sources is present within a set of ratings or scores. Two types of reliability were examined: interrater reliability, or the consistency of ratings of the same schools, and interitem reliability, or the consistency of ratings for indicators intended to reflect the same construct (see box 2).

Ratings of the same schools by school improvement coaches were reliable, even after accounting for chance. For the overall rubric the rate of agreement between school improvement coaches was .88, which is generally classified as good reliability (Altman, 1991; figure 1). The rates of agreement for each of the six components ranged from .87 (balanced assessment and collaboration) to .93 (data-based decisionmaking). An alternative consistency statistic that accounts for ratings that may occur by chance (Cohen's kappa) also indicated good interrater reliability, with kappas of .74 for the overall rubric and ranging from .71 (for multitiered instruction) to .85 (for data-based decisionmaking) on the six components (all exceeding the .60 benchmark for good reliability for kappa when accounting for chance; Landis & Koch, 1977).

For the overall rubric the rate of agreement between school improvement coaches was .88, which is generally classified as good reliability

Interitem consistency was high for the overall rubric and in the acceptable-to-good range for the six components. In addition to assessing the consistency of ratings among the district's school improvement coaches, the study also calculated the internal consistency of ratings for all indicators in the rubric and for each component (figure 2). Cronbach's

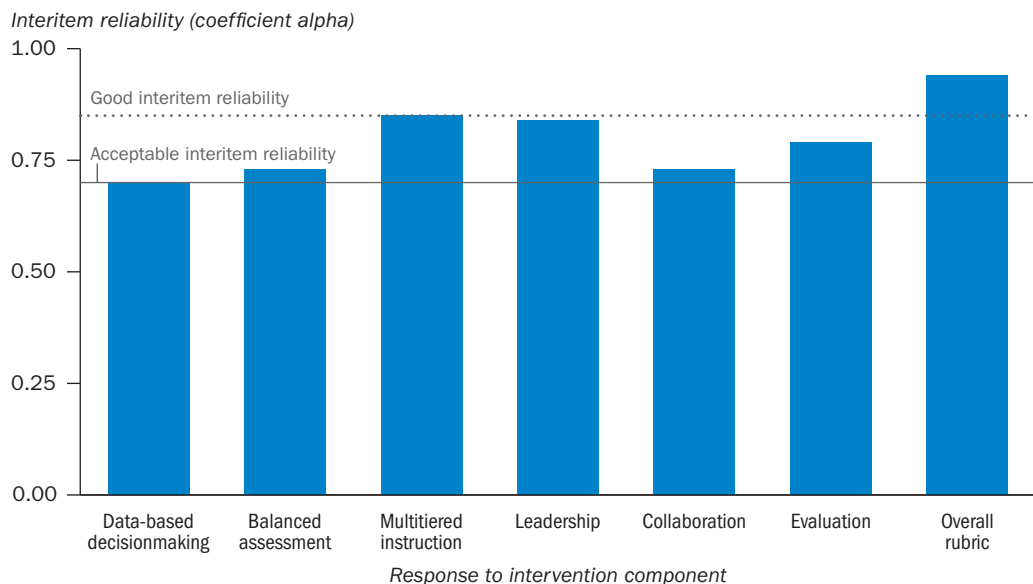
Figure 1. Interrater reliability was good for the entire implementation fidelity monitoring rubric for the Response to Intervention framework and for each component for the sample of Milwaukee Public Schools, 2014/15



Note: Sample consisted of 68 schools serving students in grades K–5. The benchmark for adequate agreement is based on Altman (1991). The benchmark for adequate reliability uses Cohen's kappa statistic, a consistency statistic that accounts for ratings that may occur by chance.

Source: Authors' analysis of implementation fidelity ratings made by Milwaukee Public Schools staff in 2014/15.

Figure 2. Interitem reliability met or exceeded the benchmark for adequate reliability for the entire implementation fidelity monitoring rubric for the Response to Intervention framework and for each component for the sample of Milwaukee Public Schools, 2014/15



Note: Sample consisted of 68 schools serving students in grades K–5.

Source: Authors' analysis of implementation fidelity ratings made by Milwaukee Public Schools staff in 2014/15.

alpha for the 33-item rubric was .94 (exceeding the benchmark of .70 for adequate reliability set for the study, based on benchmarks in George & Mallery, 2003). Alphas for the six components ranged from .70 (for data-based decisionmaking, typically considered acceptable reliability) to .85 (for multitiered instruction, typically considered good reliability).

Some 53 percent of schools visited showed adequate implementation fidelity for the Response to Intervention framework

Two types of statistics were used to judge implementation fidelity: the average of the reconciled ratings for each component and for the rubric as a whole, and the percentage of schools in each fidelity category (see box 2 for the cutpoints used to classify implementation fidelity). Component and overall averages are most useful in making fine-tuned judgments or assessing progress over time. The percentage of schools in each fidelity category is most useful as a single snapshot of how well schools are implementing RTI.

Across all schools the average ratings for the overall rubric exceeded the cutpoint for adequate implementation fidelity

At the district level average implementation ratings suggest adequate fidelity. Across all schools the average ratings for the overall rubric exceeded the cutpoint for adequate implementation fidelity (table 1). However, that average does not imply that all schools were implementing the RTI framework with adequate fidelity. Indeed, 32 schools (47 percent) were not (figure 3). The average ratings also suggest that relatively large percentages of schools had yet to adequately implement the multitiered instruction component (69 percent; 47 schools) or the evaluation component (49 percent; 33 schools), showing inadequate fidelity for these two components.

A smaller percentage of priority schools than of focus schools and other schools were implementing Response to Intervention adequately. Among school types 68 percent of priority schools had yet to attain adequate implementation fidelity for the RTI framework overall (compared with 42 percent of focus schools and 32 percent of other schools; figure 4). Across the components of the implementation fidelity rubric 82 percent of priority schools had not implemented the multitiered instruction component with adequate fidelity, and 59 percent had not implemented the evaluation component with adequate fidelity. But most priority schools (82 percent) were implementing data-based decisionmaking with adequate fidelity.

Among the subcomponents of multitiered instruction, most schools had not implemented tier 3 prevention or appropriate instruction for culturally and linguistically diverse students with fidelity

To gain more insight into the lack of adequate progress in multitiered instruction, the study team examined schools' implementation of the four subcomponents of multitiered instruction: tier 1 core curriculum, tier 2 prevention, tier 3 prevention, and culturally and linguistically responsive instruction. Across school types sample schools demonstrated little progress in implementing tier 3; 68 percent of priority schools, 56 percent of focus schools, and 63 percent of other schools were classified as showing little fidelity (figure 5). Nor had schools made much progress in implementing instruction appropriate for culturally and linguistically diverse students; 32 percent of priority schools, 21 percent of focus schools, and 22 percent of other schools were classified as demonstrating little fidelity.

Table 1. Average implementation fidelity ratings for the Response to Intervention framework for the Milwaukee Public Schools sample, by school type, 2014/15

Component	All schools			Priority schools			Focus schools			Other schools		
	Range	Mean	Standard deviation	Range	Mean	Standard deviation	Range	Mean	Standard deviation	Range	Mean	Standard deviation
Data-based decisionmaking	2.50–5.00	4.12	0.66	3.00–5.00	3.98	0.66	2.50–5.00	4.12	0.73	3.00–5.00	4.27	0.57
Balanced assessment	2.67–4.67	3.75	0.48	2.67–4.50	3.61*	0.43	2.67–4.67	3.73	0.43	2.67–4.50	3.90*	0.51
Multitiered instruction	2.07–4.87	3.32†	0.59	2.33–4.33	3.12*†	0.55	2.07–4.33	3.30†	0.56	2.27–4.87	3.53*	0.61
Leadership	2.67–5.00	4.05	0.54	3.17–4.83	3.91	0.48	3.17–5.00	4.06	0.54	2.67–5.00	4.17	0.59
Collaboration	2.33–5.00	3.81	0.57	2.67–4.33	3.70	0.50	2.33–5.00	3.78	0.63	2.67–4.67	3.95	0.55
Evaluation	2.00–5.00	3.25†	0.73	2.00–4.00	3.09†	0.65	2.00–4.50	3.08*†	0.70	2.00–5.00	3.59*	0.73
Overall rubric	2.41–4.82	3.61	0.49	2.88–4.47	3.45*†	0.42	2.41–4.38	3.59	0.48	2.47–4.82	3.79*	0.51

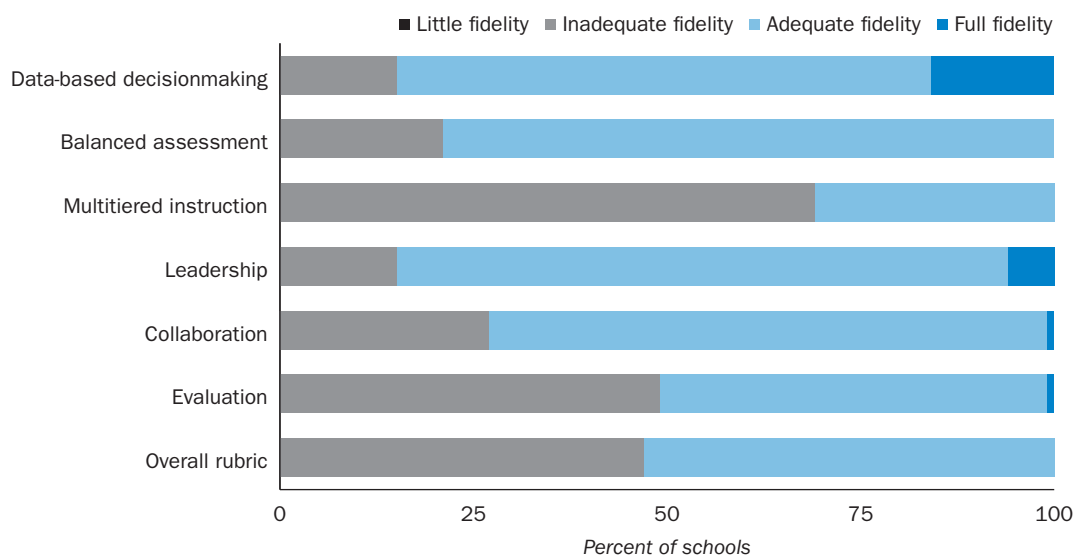
* Difference in means across school types (rows) is significant at $p < .05$.

† Average rating falls below the 3.5 cutpoint for adequate fidelity set by the National Center on Response to Intervention.

Note: The sample consists of 68 schools (22 priority schools, 24 focus schools, and 22 other schools) serving students in grades K–5. Priority schools are Title I schools in which overall student achievement was in the lowest 5 percent of Title I schools in the state. Focus schools are Title I schools in which overall student achievement was in the lowest 10 percent of Title I schools and either subgroup performance was very low or achievement gaps between subgroups were the most significant. Other schools either were not Title I schools or were not classified as priority or focus schools.

Source: Authors' analysis of data collected by Milwaukee Public Schools staff in 2014/15.

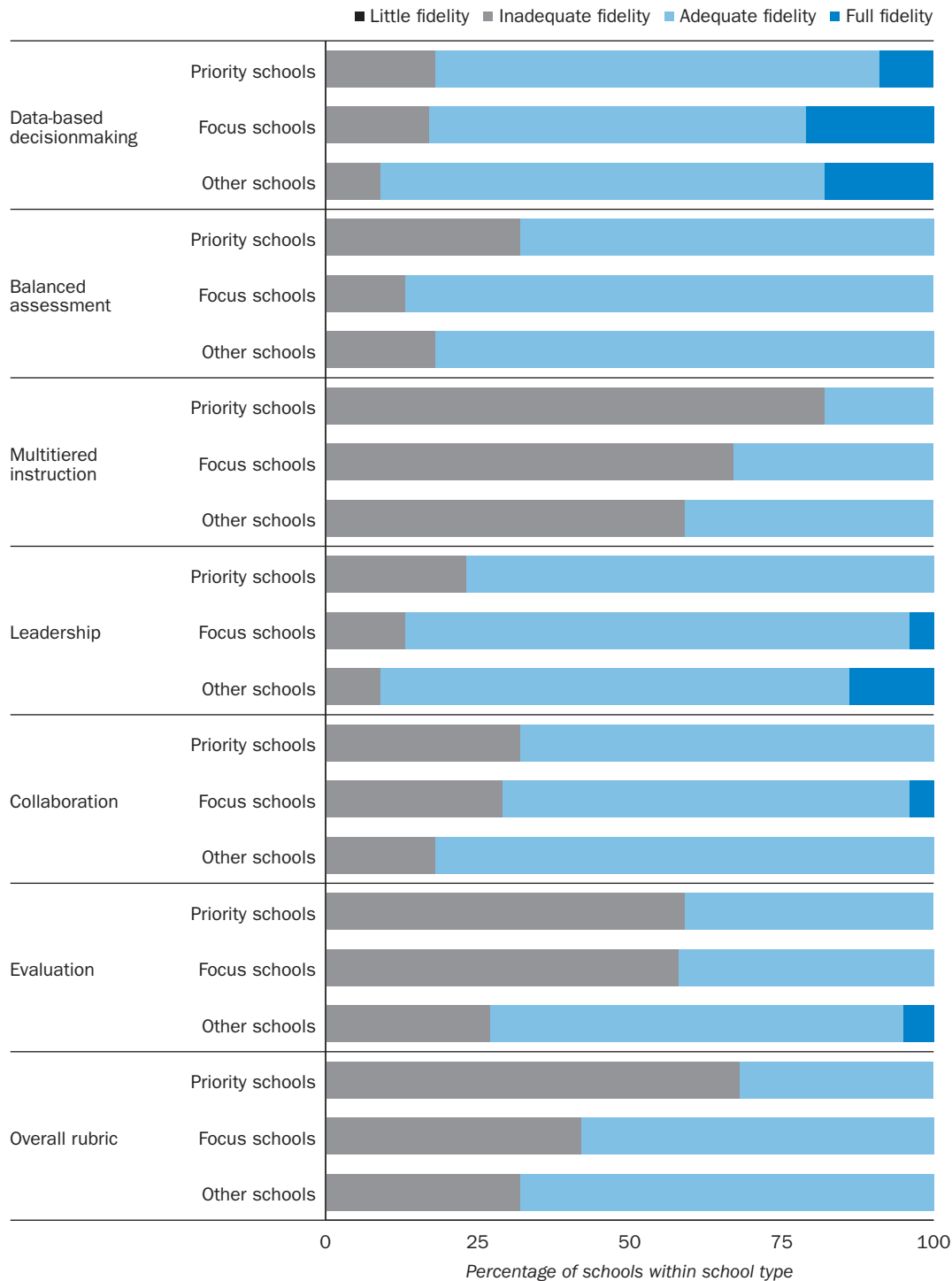
Figure 3. More than half the Milwaukee Public Schools sample had adequate implementation fidelity overall for the Response to Intervention framework, 2014/15



Note: The sample consists of 68 schools serving students in grades K–5. The cutpoints for categories of implementation fidelity are based on those used by the National Center on Response to Intervention. No components were implemented with little fidelity.

Source: Authors' analysis of implementation fidelity ratings made by Milwaukee Public Schools staff in 2014/15.

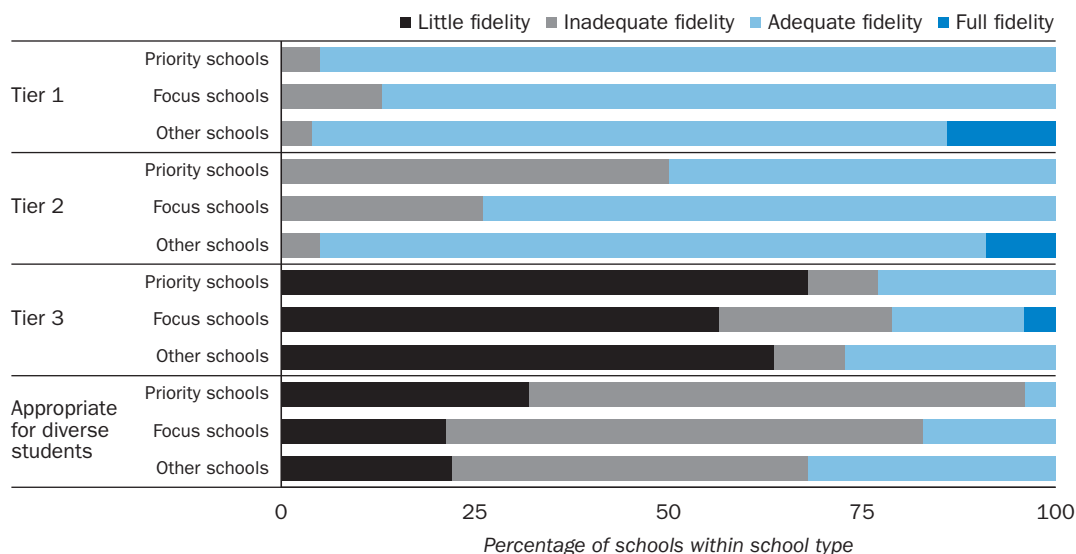
Figure 4. Adequacy of implementation fidelity for the Response to Intervention framework in the Milwaukee Public Schools sample was lowest among priority schools, followed by focus schools and other schools, 2014/15



Note: The sample consists of 68 schools (22 priority schools, 24 focus schools, and 22 other schools) serving students in grades K–5. Priority schools are Title I schools in which overall student achievement was in the lowest 5 percent of Title I schools in the state. Focus schools are Title I schools in which overall student achievement was in the lowest 10 percent of Title I schools and either subgroup performance was very low or achievement gaps between subgroups were the most significant. Other schools either were not Title I schools or were not classified as priority or focus schools. No components were implemented with little fidelity.

Source: Authors' analysis of ratings made by Milwaukee Public Schools staff in 2014/15.

Figure 5. Most schools in the Milwaukee Public Schools sample had not implemented tier 3 instruction or appropriate instruction for culturally and linguistically diverse students with fidelity, 2014/15



Note: Tier 1 involves core classroom instruction for all students, tier 2 provides supplemental instruction for students who perform poorly on subject-matter screening assessments, and tier 3 involves more individualized and intensive instruction for students who do not respond to tiers 1 and 2. The sample consists of 68 schools (22 priority schools, 24 focus schools, and 22 other schools) serving students in grades K–5. Priority schools are Title I schools in which overall student achievement was in the lowest 5 percent of Title I schools in the state. Focus schools are Title I schools in which overall student achievement was in the lowest 10 percent of Title I schools and either subgroup performance was very low or achievement gaps between subgroups were the most significant. Other schools either were not Title I schools or were not classified as priority or focus schools.

Source: Authors’ analysis of ratings made by Milwaukee Public Schools staff in 2014/15.

Implementation ratings for the Response to Intervention framework were related to characteristics of school teacher and student populations

The study team next examined the statistical relationships between school-level characteristics of students and teachers and implementation ratings.⁴ These findings could support the validity of the implementation fidelity monitoring system.

Schools with higher percentages of teachers with advanced credentials and schools with higher teacher retention rates were better at implementing RTI (table 2). In addition, schools serving more challenging student populations (schools with greater percentages of economically disadvantaged students, higher suspension rates, and lower academic achievement) had lower implementation ratings for the RTI framework. Both findings identified factors that are related to the levels of implementation fidelity.

Schools with larger percentages of teachers with advanced credentials showed stronger implementation fidelity. The correlations between teacher characteristics and implementation fidelity ratings suggest that schools with higher percentages of teachers with advanced credentials (such as a master’s degree or higher or National Board Certification) had stronger implementation fidelity for the RTI framework than did schools with lower percentages of teachers with advanced credentials. The correlations between the percentage of teachers with advanced credentials and ratings for data-based decisionmaking,

The correlations between teacher characteristics and implementation fidelity ratings suggest that schools with higher percentages of teachers with advanced credentials had stronger implementation fidelity for the RTI framework than did schools with lower percentages of teachers with advanced credentials

Table 2. Correlations between implementation fidelity ratings for the Response to Intervention framework and contextual factors for Milwaukee Public Schools, 2014/15 (Pearson product-moment correlation coefficients)

School factor	Response to Intervention component						Overall score
	Data based decision making	Balanced assessment	Multitiered instruction	Leadership	Collaboration	Evaluation	
Teacher characteristics							
Percentage of teachers with an advanced credential ^a	.28*	.22	.39***	.24	.25*	.20	.36**
Percentage of teachers with five or more years of teaching experience	.04	.20	.10	-.04	.06	-.04	.09
Percentage of teachers meeting federal highly qualified teacher requirements	-.05	.07	.12	.07	.07	-.08	.09
Student-licensed staff ratio	-.02	.10	.11	.09	.10	.18	.12
Teacher retention	.31*	.40***	.29*	.27*	.13	.11	.33**
Student characteristics							
Percentage of students who are English learner students	.17	.23	.15	-.03	-.10	.02	.12
Percentage of students with a disability	-.08	-.21	-.13	-.14	-.15	-.15	-.17
Percentage of students eligible for the federal school lunch program	-.11	-.28*	-.23	-.26*	-.21	-.26**	-.28*
Student enrollment	-.04	.02	.03	-.06	-.01	-.04	-.03
Other factors							
Percentage of students proficient in math	.21	.35**	.33**	.31**	.23	.36**	.37**
Percentage of students proficient in reading	.11	.29*	.20	.25*	.21	.23	.26*
Percentage of students suspended	-.23	-.27*	-.25*	-.25*	-.27*	-.18	-.32*

* Significant at $p < .05$; ** significant at $p < .01$; *** significant at $p < .001$.

Note: Sample consisted of 68 schools serving students in grades K–5. Benjamini–Hochberg corrections for multiple comparisons did not affect determinations of statistical significance of the correlations (Benjamini & Hochberg, 1995).

Source: Authors' analysis of ratings by school improvement coaches employed by Milwaukee Public Schools in 2014/15 and school data from the Wisconsin Department of Public Instruction website.

multitiered instruction, collaboration, and the overall implementation fidelity rubric were statistically significant.

Schools with higher teacher retention rates showed stronger implementation. Teacher retention rates between 2013/14 and 2014/15 (the year of the school visits) were positively related to schools' implementation of the RTI framework. Schools with higher teacher retention rates had higher average ratings for the overall implementation fidelity rubric and on the data-based decisionmaking, balanced assessment, multitiered instruction, and leadership components (see table 2).

Schools serving higher percentages of economically disadvantaged students had lower implementation fidelity. Schools with higher concentrations of economically disadvantaged students (defined as students eligible for the federal school lunch program) had lower

average ratings for the overall RTI implementation fidelity rubric as well as on the balanced assessment, leadership, and evaluation components.

Schools with higher student academic proficiency rates and lower suspension rates showed stronger implementation. School-level proficiency rates in math and reading were related to average implementation ratings for the overall rubric and for the balanced assessment, multitiered instruction, leadership, and evaluation components.⁵ Schools' suspension rates were negatively associated with scores for the overall rubric as well as for the balanced assessment, multitiered instruction, leadership, and collaboration components.

Implications of the study findings

The reliance of states, school districts, and schools on RTI as a school improvement strategy (as indicated in states' Elementary and Secondary Education Act waivers) reflects the belief that RTI can help meet students' instructional needs and increase the learning growth of students who are falling behind their peers academically. If RTI is indeed an effective school improvement strategy, using an implementation fidelity monitoring system such as that described in this report may improve the chances that RTI will produce the expected impacts.

The study found that the implementation fidelity monitoring system for the RTI framework can produce reliable evidence. The consistency of the school improvement coaches' ratings (interrater reliability) was adequate even when the analysis accounted for chance ratings. In addition, the indicators that make up the rubric and the components of the RTI framework showed good internal consistency. The reliability findings support the use of the implementation monitoring rubric in other schools and districts that are attempting to establish RTI.

For Milwaukee Public Schools the results generated by this implementation fidelity monitoring system have informed school and school district administrators about which RTI processes need the most improvement. Overall, 53 percent of schools were implementing the RTI framework with adequate fidelity. Across school types, schools were implementing the leadership and collaboration components with the strongest level of implementation fidelity; schools have made the least amount of progress in implementing the multitiered instruction and evaluation components. Looking deeply into schools' implementation of multitiered instruction, it appears that schools can benefit most from professional development on tier 3 instruction and culturally and linguistically responsive instruction.

The implementation fidelity findings suggest that priority schools need the most professional development and coaching support on RTI, followed by focus schools. Schools classified as other (that is, schools that are not Title I schools and schools that are better performing) need the least amount of professional development on RTI.

This report also describes a process that other districts can follow to develop their own implementation fidelity monitoring system for the RTI framework (see appendix A) or for other interventions. RTI represents a process that schools and districts can use to target instructional resources to students requiring additional support. Fidelity monitoring systems perform a similar function: they can target professional development activities to the components of interventions that need most improvement.

It appears that schools can benefit most from professional development on tier 3 instruction and culturally and linguistically responsive instruction

Limitations of the study

The primary limitation of the study is its lack of generalizability. Although the school improvement coaches were consistent in rating schools' implementation of RTI, they visited schools in only one urban district. Moreover, the schools were not randomly selected. Rather, school principals volunteered to participate. The study originated as a need shared by several districts participating in the Midwest Urban Research Alliance. Their adoption of the RTI implementation fidelity monitoring system should provide information on whether the system can be installed and can prove reliable in other settings.

Another limitation was the inability to obtain in-depth information on school staff perceptions of the system or of the findings it produces. In addition, the study had to rely on publicly available school-level information, which meant that information on other teacher and student characteristics that might have played an influential role in schools' implementation fidelity was not examined.

Also, the correlations between school characteristics and implementation ratings (see table 2) provide no information about the causal direction between variables. While these correlations may indicate that the identified school characteristics facilitate or hinder schools' ability to implement RTI, it might instead be the case that success in implementing RTI is resulting in better school-level characteristics, such as academic outcomes, or that both implementation scores and school outcomes are a result of other, unmeasured variables. Future research can investigate the direction of causality using an appropriate research design.

Finally, the study did not investigate whether monitoring the implementation fidelity of the RTI framework improves student outcomes. Although it is reasonable to expect implementation monitoring to lead to necessary modifications and thereby improve student outcomes, the study was not designed to address this question. Future, more rigorous studies could investigate this possibility.

The study did not investigate whether monitoring the implementation fidelity of the RTI framework improves student outcomes

Appendix A. The district's implementation fidelity monitoring system for the Response to Intervention framework

This appendix describes the collaboration process for developing the rubric and system for monitoring implementation fidelity of the RTI framework and provides the contents of the rubric.

Collaborative process for developing the rubric and system

During the 2013/14 school year Regional Educational Laboratory (REL) Midwest and its partners at the National Center on Response to Intervention collaborated with administrators and staff from Milwaukee Public Schools to develop the district's system for monitoring the implementation fidelity of the RTI framework.

REL Midwest's technical assistance team developed the rubric and system based on feedback from potential system users. The process involved the following six steps:

- 1. Interviews with potential users.** The technical assistance team conducted interviews with nine potential users about the strengths and weaknesses of the rubrics they were using at the time, types of information they were seeking about implementation of RTI, their preferred format for seeing results, and their preferred system features. These interviews were conducted with district administrators, school improvement coaches, school administrators, and members of schools' RTI implementation teams. From these interviews, the study team identified the following themes:
 - District administrators requested that any rubric include items found in a self-administered fidelity rubric already in use and items from the district's improvement plan.
 - School administrators emphasized that the rubric and system must provide information that is useful, easy to access, and easy to interpret.
 - School RTI implementation team members emphasized that an implementation fidelity monitoring system must be used for formative purposes only (for continuous improvement).

- 2. Development of a hybrid rubric.** REL Midwest's technical assistance team then did an indicator-by-indicator comparison of three documents: the Response to Intervention Essential Components Integrity Rubric of the National Center on Response to Intervention (consisting of 26 indicators; National Center on Response to Intervention, 2010), the School-wide Implementation Review (a self-assessment completed annually by schools to measure RTI processes consisting of 61 indicators), and the 30 indicators included in the district's improvement plan. Duplicate indicators were removed or consolidated, yielding a hybrid rubric consisting of 33 indicators grouped into six components.

- 3. Hybrid rubric and dashboard prototype presented to focus group.** The REL Midwest technical assistance team developed a prototype of a data dashboard showing how results can be aggregated and disaggregated, along with descriptors for ratings of 1, 3, and 5 for each rubric indicator. Focus group feedback was integrated into follow-on versions of the rubric and prototype dashboard.

4. ***Training materials for district staff presented to a second focus group.*** The revised rubric and dashboard prototype, along with training materials on how to use them, were presented to another focus group for comment. Focus group members suggested a few changes but were unable to provide much feedback on the training materials before district staff underwent training and attempted to use the system to make ratings.
5. ***Training of district staff to gather evidence at schools and make ratings.*** The training materials were piloted with a group of eight Milwaukee Public Schools staff. Once these staff rated training scenarios that were consistent with those of a master rater, they made site visits to four schools to test their ability to generate reliable and valid ratings in the field and to identify gaps in the training.
6. ***System refinement based on district staff comments.*** The district school improvement coaches who made the school visits offered one last round of feedback to the technical assistance team, based on their hands-on experiences with using the rubric and system in four schools (two coaches rated each of the four schools).

Contents of the rubric

Below is a list of the components, subcomponents, and indicators of the implementation fidelity rubric. The components are listed in black, the subcomponents in blue, and the indicators in grey. Table A1 describes these elements in detail, along with the descriptions for ratings 1, 3, and 5.

Data-based decisionmaking

- Decisionmaking process
- Data system

Balanced assessment

Screening

- Screening tools
- Universal screening
- Multiple data points

Progress monitoring

- Progress-monitoring tools
- Monitoring progress

Culturally and linguistically responsive assessment

- Appropriate assessments for culturally and linguistically diverse students

Multitiered instruction

Tier 1 core curriculum

- Research-based curriculum materials
- Articulation of teaching and learning (in and across grade levels)
- Instruction
- Standards-based curriculum

Tier 2 prevention

- Evidence-based intervention
- Complements core instruction
- Instruction
- Determining responsiveness to tier 2

Addition to tier 1

Tier 3 prevention

Data-based interventions adapted based on student need

Instruction

Determining responsiveness to tier 3

Relationship to tier 1

Culturally and linguistically responsive instruction

Appropriate instruction for culturally and linguistically diverse students

Leadership

Response to Intervention focus

Leadership

Staff

School-based professional development

Schedules

Resources

Collaboration

Communication with and involvement of parents

Communication with and involvement of all staff

Response to Intervention teams

Evaluation

Evaluation

Fidelity

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Component 1: Data-based decisionmaking—Data-based decisionmaking processes are used to inform instruction, movement within the multitiered system, and disability identification (in accordance with state law).			
Decisionmaking process	The mechanism for making decisions about the participation of students across tiers meets no more than one of the following criteria: the process (1) is data driven and based on validated methods; (2) involves a broad base of stakeholders; or (3) is operationalized with clear, established decision rules (such as movement between tiers and determination of appropriate instruction or interventions).	The mechanism for making decisions about the participation of students across tiers meets two of these criteria: the process (1) is data driven and based on validated methods; (2) involves a broad base of stakeholders; or (3) is operationalized with clear, established decision rules (such as movement between tiers and determination of appropriate instruction or interventions).	The mechanism for making decisions about the participation of students across tiers meets all of these criteria: the process (1) is data driven and based on validated methods; (2) involves a broad base of stakeholders; and (3) is operationalized with clear, established decision rules (such as movement between tiers and determination of appropriate instruction or interventions).
Data system	No data system is in place to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions.	A data system is partially in place to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions.	A comprehensive data system is in place to document and access individual student-level data (including screening and progress-monitoring data) and instructional decisions.

(continued)

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5 (continued)

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Component 2: Balanced assessment—Screening, progress monitoring, and other supporting assessments are used to inform decisionmaking.			
<i>Subcomponent 1: Screening—The Response to Intervention framework accurately identifies students at risk for poor learning outcomes or challenging behaviors.</i>			
Screening tools	There is insufficient evidence that the screening tools are reliable, or that correlations between the instruments and valued outcomes are strong, or that predictions of risk status are accurate.	Evidence indicates that the screening tools are reliable and that correlations between the instruments and valued outcomes are strong. However, there is insufficient evidence that predictions of risk status are accurate, and staff may be unable to articulate the supporting evidence.	Evidence indicates that the screening tools are reliable, correlations between the instruments and valued outcomes are strong, predictions of risk status are accurate, and staff is able to articulate the supporting evidence.
Universal screening	No conditions are met: (1) Screening is conducted for all students (is universal); (2) procedures are in place to ensure implementation accuracy (all students are tested, scores are accurate, cutpoints are accurate); or (3) screening of all students occurs more than once per year (such as fall, winter, and spring).	One or two conditions are met: (1) Screening is conducted for all students (is universal); (2) procedures are in place to ensure implementation accuracy (all students are tested, scores are accurate, cutpoints are accurate); or (3) screening of all students occurs more than once per year (such as fall, winter, and spring).	All conditions are met: (1) Screening is conducted for all students (is universal); (2) procedures are in place to ensure implementation accuracy (all students are tested, scores are accurate, cutpoints are accurate); and (3) screening of all students occurs more than once per year (such as fall, winter, and spring).
Multiple data points	Screening data are used alone to decide whether a student is at risk.	Screening data are used in concert with at least one other type of data (such as classroom performance, curriculum-based assessment, performance on state assessments, diagnostic assessment data, or short-term progress monitoring) to decide whether a student is at risk.	Screening data are used in concert with at least two other types of data (such as classroom performance, curriculum-based assessment, performance on state assessments, diagnostic assessment data, or short-term progress monitoring) to decide whether a student is at risk.
<i>Subcomponent 2: Progress monitoring—Ongoing and frequent monitoring of progress quantifies rates of improvement and informs instructional practice and the development of individualized programs.</i>			
Progress-monitoring tools	Selected progress-monitoring tools meet no more than one of the following criteria: (1) has at least nine alternate forms of equal and controlled difficulty; (2) specifies minimum acceptable growth; (3) provides benchmarks for minimum acceptable end-of-year performance; or (4) provides reliability and validity information for the performance-level score.	Selected progress-monitoring tools meet two or three of the following criteria: (1) has at least nine alternate forms of equal and controlled difficulty; (2) specifies minimum acceptable growth; (3) provides benchmarks for minimum acceptable end-of-year performance; or (4) provides reliability and validity information for the performance-level score.	Selected progress-monitoring tools meet all of the following criteria: (1) has at least nine alternate forms of equal and controlled difficulty; (2) specifies minimum acceptable growth; (3) provides benchmarks for minimum acceptable end-of-year performance; and (4) provides reliability and validity information for the performance-level score and enables staff to articulate the supporting evidence.

(continued)

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5 (continued)

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Monitoring progress	Neither condition is met: (1) Progress monitoring occurs at least monthly for students receiving tier 2 interventions and at least weekly for students receiving tertiary interventions; or (2) procedures are in place to ensure implementation accuracy (appropriate students are tested; scores are accurate; decisionmaking rules are applied consistently).	Only one condition is met: (1) Progress monitoring occurs at least monthly for students receiving tier 2 interventions and at least weekly for students receiving tertiary interventions; or (2) procedures are in place to ensure implementation accuracy (appropriate students are tested; scores are accurate; decisionmaking rules are applied consistently).	Both conditions are met: (1) Progress monitoring occurs at least monthly for students receiving tier 2 interventions and at least weekly for students receiving tertiary interventions; and (2) procedures are in place to ensure implementation accuracy (appropriate students are tested; scores are accurate; decisionmaking rules are applied consistently).
<i>Subcomponent 3: Culturally and linguistically responsive assessment</i>			
Appropriate assessments for culturally and linguistically diverse students	Assessments do not account for cultural, linguistic, and socioeconomic factors.	Assessments strive to consider cultural, linguistic, and socioeconomic factors, but some areas need improvement.	Assessments reflect cultural, linguistic, and socioeconomic factors.
Component 3: Multitiered instruction—The framework includes a schoolwide, multitiered system for preventing school failure.			
<i>Subcomponent 1: Tier 1 core curriculum</i>			
Research-based curriculum materials	The core curriculum materials largely are not research-based for the target population of learners (including subgroups).	Some of the core curriculum materials are research-based for the target population of learners (including subgroups).	All of the core curriculum materials are research-based for the target population of learners (including subgroups).
Articulation of teaching and learning (in and across grade levels)	Neither condition is met: (1) Teaching and learning are well articulated from one grade to another; or (2) teaching and learning are well articulated within grade levels so students have highly similar experiences, regardless of their assigned teacher.	Only one condition is met: (1) Teaching and learning are well articulated from one grade to another; or (2) teaching and learning are well articulated within grade levels so students have highly similar experiences, regardless of their assigned teacher.	Both conditions are met: (1) Teaching and learning are well articulated from one grade to another; and (2) teaching and learning are well articulated within grade levels so students have highly similar experiences, regardless of their assigned teacher.
Instruction	Neither condition is met: (1) Most or all teachers differentiate instruction; or (2) teachers use students' assessment data to identify students' needs.	Only one condition is met: (1) Most or all teachers differentiate instruction; or (2) teachers use students' assessment data to identify students' needs.	Both conditions are met: (1) Most or all teachers differentiate instruction; and (2) teachers use students' assessment data to identify students' needs.
Standards-based curriculum	The universal curriculum and instruction are not aligned with Common Core State Standards.	The universal curriculum and instruction are partially aligned with Common Core State Standards.	All universal curriculum and instruction are aligned with Common Core State Standards.
<i>Subcomponent 2: Tier 2 prevention</i>			
Evidence-based intervention	Tier 2 interventions are not evidence based.	Tier 2 interventions consist of a variety of strategies, of which only some are evidence based and some are not.	All tier 2 interventions are evidence based.
Complements core instruction	Tier 2 is poorly aligned with core instruction and incorporates different topics, even though those topics are not foundational skills that support core instruction.	Tier 2 is generally aligned with core instruction but only occasionally incorporates foundational skills that support core instruction.	Tier 2 is well aligned with core instruction and incorporates foundational skills that support core instruction.

(continued)

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5 (continued)

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Instruction	Neither condition is met: (1) Tier 2 interventions are led by staff trained in the intervention according to developer requirements; or (2) group size is optimal (according to research) for the age and needs of students.	Only one condition is met: (1) Tier 2 interventions are led by staff trained in the intervention according to developer requirements; or (2) group size is optimal (according to research) for the age and needs of students.	Both conditions are met: (1) Tier 2 interventions are led by staff trained in the intervention according to developer requirements; and (2) group size is optimal (according to research) for the age and needs of students.
Determining responsiveness to tier 2	Neither condition is met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement or end-of-year benchmarks; or (2) these decisionmaking criteria are implemented accurately.	Only one condition is met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement or end-of-year benchmarks; or (2) these decisionmaking criteria are implemented accurately.	Both conditions are met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement or end-of-year benchmarks; and (2) these decisionmaking criteria are implemented accurately.
Addition to tier 1	Tier 2 interventions replace core instruction.	Tier 2 interventions sometimes supplement core instruction and sometimes replace core instruction.	Tier 2 interventions supplement core instruction.
<i>Subcomponent 3: Tier 3 prevention</i>			
Data-based interventions adapted based on student need	Tier 3 interventions are not more intensive than tier 2 interventions (for example, no increase in duration or frequency, change in the interventionist, change in group size, or change in type of intervention).	Tier 3 interventions are more intensive than tier 2 interventions based only on preset methods to increase intensity (for example, sole reliance on increased duration or frequency, a change in the interventionist, a smaller group size, and a change in type of intervention).	Tier 3 interventions are more intensive than tier 2 interventions and are adapted to address individual student needs in a number of ways through an iterative manner based on student data (for example, increased duration or frequency, a change in the interventionist, a smaller group size, a change in instructional delivery, and a change in type of intervention).
Instruction	Neither condition is met: (1) Tier 3 interventions are led by well-trained staff experienced in individualizing instruction based on data; or (2) group size is optimal (according to research) for the age and needs of students.	Only one condition is met: (1) Tier 3 interventions are led by well-trained staff experienced in individualizing instruction based on data; or (2) group size is optimal (according to research) for the age and needs of students.	Both conditions are met: (1) Tier 3 interventions are led by well-trained staff experienced in individualizing instruction based on data; and (2) group size is optimal (according to research) for the age and needs of students.
Determining responsiveness to tier 3	Neither condition is met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement, end-of-year benchmarks, or an intra-individual framework; or (2) these decisionmaking criteria are implemented accurately.	Only one condition is met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement, end-of-year benchmarks, or an intra-individual framework; or (2) these decisionmaking criteria are implemented accurately.	Both conditions are met: (1) Decisions about responsiveness to intervention are based on reliable and valid progress-monitoring data to reflect slope of improvement, end-of-year benchmarks, or an intra-individual framework; and (2) these decisionmaking criteria are implemented accurately.

(continued)

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5 (continued)

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Relationship to tier 1	Neither condition is met: (1) Decisions regarding student participation in both primary and tertiary levels of prevention are made on a case-by-case basis, according to student need; or (2) tertiary-level interventions address the general education curriculum in an appropriate manner for students.	Only one condition is met: (1) Decisions regarding student participation in both primary and tertiary levels of prevention are made on a case-by-case basis, according to student need; or (2) tertiary-level interventions address the general education curriculum in an appropriate manner for students.	Both conditions are met: (1) Decisions regarding student participation in both primary and tertiary levels of prevention are made on a case-by-case basis, according to student need; and (2) tertiary-level interventions address the general education curriculum in an appropriate manner for students.
<i>Subcomponent 4: Culturally and linguistically responsive instruction</i>			
Appropriate instruction for culturally and linguistically diverse students	Instruction and interventions do not account for cultural, linguistic, and socioeconomic factors.	Instruction and interventions strive to consider cultural, linguistic, and socioeconomic factors, but some areas need improvement.	Instruction and interventions reflect cultural, linguistic, and socioeconomic factors.
Component 4: Leadership			
Response to Intervention focus	Staff perceives Response to Intervention as a pre-referral process that students must complete to be referred to special education.	Differences are noted among staff regarding their understanding of the purpose of Response to Intervention.	Staff believes that the primary purpose of Response to Intervention is to support all students based on need, including providing early interventions to prevent students from having academic and behavioral problems and enriching opportunities for students exceeding benchmarks.
Leadership	Decisions and actions by school and district leaders undermine the effectiveness of the essential components of the Response to Intervention framework at the school.	Decisions and actions by school and district leaders are inconsistent and only somewhat supportive of the essential components of the Response to Intervention framework at the school.	Decisions and actions by school and district leaders proactively support the essential components of the Response to Intervention framework at the school and make the Response to Intervention framework more effective.
Staff	Neither condition is met: (1) Staff are highly qualified to deliver interventions and instruction for students at all tiers and are adequately trained for their responsibilities; or (2) staff are allocated to support the delivery of multitiered instruction based on student need.	Only one condition is met: (1) Staff are highly qualified to deliver interventions and instruction for students at all tiers and are adequately trained for their responsibilities; or (2) staff are allocated to support the delivery of multitiered instruction based on student need.	Both conditions are met: (1) Staff are highly qualified to deliver interventions and instruction for students at all tiers and are adequately trained for their responsibilities; and (2) staff are allocated to support the delivery of multitiered instruction based on student need.
School-based professional development	The school has no well-defined, school-based professional development mechanism to support continuous improvement of instructional practice.	Some forms of professional development are available to teachers to support continuous improvement of instructional practice, but most are not school-based and do not establish a means of continuously improving instructional practice.	School-based professional development is institutionalized and structured so that all teachers continuously examine, reflect upon, and improve instructional practice.
Schedules	Schoolwide schedules are not aligned to support multiple tiers of prevention and high-quality instruction based on student need.	Schoolwide schedules are partially aligned to support multiple tiers of prevention and high-quality instruction based on student need.	Schoolwide schedules are aligned to support multiple tiers of prevention and high-quality instruction based on student need.

(continued)

Table A1. Final implementation fidelity rubric for the Response to Intervention framework showing components, subcomponents, and indicators and the descriptors for ratings 1, 3, and 5 (continued)

Component, subcomponent, and indicator	Rating 1	Rating 3	Rating 5
Resources	Resources (funds and programs) are not allocated to support Response to Intervention implementation.	Resources (funds and programs) are partially allocated to support Response to Intervention implementation.	Resources (funds and programs) are allocated to support Response to Intervention implementation.
Component 5: Collaboration			
Communication with and involvement of parents	No conditions are met: (1) A description of the school's essential components of Response to Intervention is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving tier 2 or Tier 3 interventions; and (3) parents are involved during decisionmaking.	One or two conditions are met: (1) A description of the school's essential components of Response to Intervention is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving tier 2 or tier 3 interventions; or (3) parents are involved during decisionmaking.	All conditions are met: (1) A description of the school's essential components of Response to Intervention is shared with parents; (2) a coherent mechanism is implemented for updating parents on the progress of their child who is receiving tier 2 or tier 3 interventions; and (3) parents are involved during decisionmaking.
Communication with and involvement of all staff	No conditions are met: (1) A description of the school's essential components of Response to Intervention and data-based decisionmaking process are shared with staff; (2) a system is in place to keep staff informed; or (3) teacher teams collaborate frequently.	One or two conditions are met: (1) A description of the school's essential components of Response to Intervention and data-based decisionmaking process are shared with staff; (2) a system is in place to keep staff informed; or (3) teacher teams collaborate frequently.	All conditions are met: (1) A description of the school's essential components of Response to Intervention and data-based decisionmaking process are shared with staff; (2) a system is in place to keep staff informed; and (3) teacher teams collaborate frequently.
Response to intervention teams	Neither condition is met: (1) The Response to Intervention team is representative of all key stakeholders; or (2) structures and clear processes are in place that enable the team to meet regularly.	Only one condition is met: (1) The Response to Intervention team is representative of all key stakeholders; or (2) structures and clear processes are in place that enable the team to meet regularly.	Both conditions are met: (1) The Response to Intervention team is representative of all key stakeholders; and (2) structures and clear processes are in place that enable the team to meet regularly.
Component 6: Evaluation			
Evaluation	No conditions are met: (1) An evaluation plan is in place to monitor short- and long-term goals; (2) student data are reviewed for all students and subgroups of students across the essential components (core curriculum is effective, interventions are effective, screening process); or (3) implementation data are reviewed to monitor fidelity and efficiency across all components of the Response to Intervention framework.	One or two conditions are met: (1) An evaluation plan is in place to monitor short- and long-term goals; (2) student data are reviewed for all students and subgroups of students across the essential components (core curriculum is effective, interventions are effective, screening process); or (3) implementation data are reviewed to monitor fidelity and efficiency across all components of the Response to Intervention framework.	All conditions are met: (1) An evaluation plan is in place to monitor short- and long-term goals; (2) student data are reviewed for all students and subgroups of students across the essential components (core curriculum is effective, interventions are effective, screening process); and (3) implementation data are reviewed to monitor fidelity and efficiency across all components of the Response to Intervention framework.
Fidelity	Neither condition is met: (1) Procedures are in place to monitor the implementation fidelity of the core curriculum, secondary and tertiary interventions, and assessments; or (2) the preponderance of evidence supports fidelity.	Only one condition is met: (1) Procedures are in place to monitor the implementation fidelity of the core curriculum, secondary and tertiary interventions, and assessments; or (2) the preponderance of evidence supports fidelity.	Both conditions are met: (1) Procedures are in place to monitor the implementation fidelity of the core curriculum, secondary and tertiary interventions, and assessments; and (2) the preponderance of evidence supports fidelity.

Appendix B. Data collection and methodology

This appendix describes the sampling strategy, data collection, and the analyses for each research question.

Sampling strategy

The target number of 70 schools represented the maximum number of schools that could be visited by the district's school improvement coaches in a single school year with available resources. For the 2014/15 school year there were 116 schools governed by Milwaukee Public Schools that served students in grades K–5, all of which had been implementing the RTI framework since 2012/13. District officials contacted the principal at each of the schools serving grades K–5 to discuss the importance to the school and district of the RTI implementation fidelity monitoring system as well as the minimum burden involved in participation. If the principal agreed to participate, the school was eligible to be in the sample. District officials were able to recruit 70 schools, so sampling was not necessary. Because of incomplete data collection from two schools, the final sample included 68 schools.

Data sources, instruments, and collection methods

Training of school improvement coaches in use of rubric and system. The 23 school improvement coaches that were eligible to make school visits participated in a three-day training led by the same former affiliates of the National Center on Response to Intervention who helped develop the implementation fidelity monitoring system for the RTI framework.⁶ Training covered an overview of RTI and its component processes, the indicators for each component, types of evidence to look for to make ratings on the indicators, and how to enter the ratings into the data dashboard. The participants had opportunities to complete five practice exercises prior to the certification test. Practice exercises provided an opportunity for participants to check their understanding by responding to scenarios related to the topics addressed on the same day. All 23 school improvement coaches who participated in the training took the certification test, and every trained school improvement coach, except one, exceeded the minimal proficiency level of 80 percent agreement. Approximately 40 percent of the school improvement coaches exceeded 90 percent proficiency; four had perfect scores.

School improvement coaches' ratings. School improvement coaches visited 70 schools and marked their ratings on the implementation fidelity monitoring system rubric on the basis of the evidence obtained during the visit. Two school improvement coaches visited each school. As specified in the implementation fidelity monitoring system, the district's school improvement coaches interviewed the principal and key members of the school RTI implementation team, including K–5 teachers and other school staff with RTI implementation duties at the school. The school improvement coaches also examined other available data that document RTI implementation in K–5 classrooms (for example, student schedules). Each school visit, including onsite data collection and data entry into the data dashboard, lasted one day. At each school the two school improvement coaches rated implementation independently of the other and did not share their ratings before the data were entered into the dashboard. The two coaches for each school compared their ratings on all 33 indicators, discussed discrepant ratings, and agreed on final reconciled ratings. One school improvement coach then entered the final set of agreed-to ratings into the dashboard.

School characteristics. The study team obtained data on school-level student and teacher characteristics from two sources: the Wisconsin Department of Public Education’s publicly accessible databases and the Milwaukee Public Schools district’s databases. The Wisconsin Department of Public Instruction’s website has student enrollment, demographic, and academic performance data stored in one site and school staff data in another site. The study team obtained implementation ratings through a formal data request to the district. Data from these sources were merged, with the name of the school and school identification number used to link the data.

The variables of interest for the study represent factors that might facilitate or impede a school’s ability to implement the RTI framework (table B1). The study team analyzed the variables of greatest relevance to district staff who were focused on implementation of RTI.

Data processing and analysis

The study was designed to understand implementation of the RTI framework by addressing questions related to the reliability of the implementation fidelity rubric, the actual levels of implementation within schools, and relationships between the implementation levels and characteristics of the schools. The REL Midwest study team analyzed the implementation indicator ratings and available school characteristics to address the research questions. Specifically, the study team calculated reliability indexes (first research question), rubric component scores (second research question), and correlation coefficients (third research question).

First research question: How reliable is the district’s implementation fidelity monitoring system for the Response to Intervention framework? An important consideration for data collected through observations is the degree of consistency in the ratings. Reliability

Table B1. Contextual factors that might be related to the implementation fidelity of the Response to Intervention framework

Teacher related factors	Student related factors	Other schoolwide factors
Teacher retention during the previous year ^a	Percentage of students who are not proficient speakers of English ^b	School performance category (priority, focus, and other) ^a
Percentage of teachers with advanced academic credentials (master’s degrees or higher or National Board Certification) ^b	Percentage of students with a disability ^b	Percentage of students proficient in math during the past three years ^{b,c}
Percentage of teachers with five or more years of teaching experience ^b	Percentage of economically disadvantaged students ^b	Percentage of students proficient in reading during the past three years ^{b,c}
Percentage of teachers meeting federal highly qualified teacher requirements ^b	Number of students enrolled ^b	Percentage of students suspended per year ^b
Pupil–teacher ratio ^b		

a. Available from Wisconsin Department of Public Instruction’s data dashboard.

b. Available from the Milwaukee Public Schools department of research.

c. Sum of the annual number of students who scored at the proficient or advanced levels as a share of the sum of the annual total number of students.

Source: Authors’ compilation.

represents the consistency in ratings or scores, or conversely, the degree to which error from various sources is present within a set of ratings or scores. To address the first research question, the study team examined two types of reliability: interrater reliability (the amount of measurement error due to differences among the raters) and interitem reliability (the amount of measurement error due to differences among indicators that assess similar constructs).

Interrater reliability. To determine the degree of consistency among raters, the study team calculated reliability in two ways. The first calculation produced the actual rate of agreement, while the second calculation accounted for chance factors in determining reliability.

The rate of agreement between the school improvement coaches is calculated using the following formula:

$$\text{Interrater reliability} = \text{number ratings with agreement} / \text{total number ratings}$$

This statistic (referred to as percent agreement) has intuitive appeal for practitioners. The developers of the customized Milwaukee Public Schools rubric (as well as the developers of the National Center on Response to Intervention rubric) considered ratings to be in agreement if the school improvement coaches who visited a school had ratings that were within one point of each other for each indicator.

However, some level of agreement could occur simply if school improvement coaches made random ratings using the rubric (referred to here as “chance agreement”). The Cohen’s kappa statistic produced another estimate of interrater reliability that accounted for chance agreement. The kappa statistic values range from -1 to 1 , where 1 indicates perfect agreement, 0 indicates chance agreement, and -1 indicates systematic disagreement between observers (Viera & Garret, 2005).

The general formula for Cohen’s kappa is:

$$K = p_o - p_e / (1 - p_e)$$

where p_o is the observed level of agreement and p_e is the expected level of agreement.

For the study the expected level of agreement was set to .60.

Interitem reliability. Interitem reliability indicates the degree to which items in a measure reflect the same underlying construct. If interitem reliability is low, practitioners cannot be certain whether the numbers reflect a single construct (for example, implementation of RTI) or a number of distinct constructs. To determine the level of interitem reliability, the study team calculated seven statistics:

- A coefficient alpha⁷ based on all 33 indicators within the rubric.
- A coefficient alpha for each of the six components of the rubric.

For the study the benchmark for adequate interitem reliability was set to .70 for the coefficient alpha and correlation coefficient for the two-item rubric component.⁸ The overall coefficient alpha for the rubric and for each of the six components met or exceeded the benchmark for reliability.

Second research question: To what extent are participating elementary schools implementing the Response to Intervention framework with fidelity? To provide direct information about the degree to which schools were implementing the RTI framework, the study team calculated average ratings for each school for each of the six components of the rubric as well as an overall implementation score. To classify schools' level of fidelity to implementation of the RTI framework, the study team adopted the cutpoints recommended by partners who had previously worked for the National Center on Response to Intervention, but with category labels appropriate for schools in a district that had rolled-out RTI implementation two years before.⁹ The fidelity categories and associated cutpoints were as follows:

- Schools with an average rating of less than 2.00 were classified as having made little progress at implementing the RTI framework (low implementation fidelity).
- Schools with an average rating of 2.00–3.49 were classified as demonstrating some progress at implementing RTI but still inadequate.
- Schools with an average rating of 3.50–4.99 were classified as having adequate implementation fidelity.
- Schools with an average rating of 5.00 were classified as demonstrating full implementation.¹⁰

Finding that the rubric and components were considered reliable, the study team continued by calculating the following:

- Descriptive statistics across all sampled schools in the district for the three performance-related groups of schools (priority, focus, and other; see box 2 in main report) and by component.
- Percentages of schools in each fidelity category, for the overall rubric and for each component.

Third research question: Are schools' average implementation ratings statistically related to school characteristics, such as teacher characteristics, student characteristics, and other schoolwide factors? To provide insight into the rubric scores, the study team calculated correlations between school contextual factors and schools' average rubric scores. These relationships were examined for two reasons. The selected contextual factors could support the validity of the fidelity measure. For instance, one would expect schools with larger percentages of students who lack basic English-speaking skills or require special education services to experience more challenges when implementing the tiered instruction component of the rubric, because a larger number of students would need more small-group or individualized instruction.

Exploring correlations between school contextual factors and implementation fidelity of the RTI framework might provide more information for school and district decision-making. The correlations could reveal factors that influence schools' success at implementing the RTI framework. For example, if high percentages of teachers in a school have a master's degree or higher, those teachers may find it easier to implement RTI because of their exposure to the RTI concept in their master's program. Power estimates indicate that the study, which involved site visits to 70 schools, had sufficient power to detect a correlation of .33 at the .05 alpha level 80 percent of the time (power at 80).

The study team examined the relationships between the overall rubric scores and contextual variables by calculating Pearson product-moment correlation coefficients. To study

relationships for categorical variables such as school region and school performance category, the study team conducted analyses of variance, with the region and performance categories serving as independent variables and average implementation fidelity ratings serving as outcome variables. If analysis of variance showed significant differences between groups, then differences between specific groups were explored.

The study team attempted to restrict the number of possible correlates to the bare minimum to reduce the chances of obtaining a significant correlation by chance. Even so, these analyses introduce an increased likelihood for false positive results by examining multiple relationships. To handle this problem, the study team performed the Benjamini–Hochberg correction post hoc (Benjamini & Hochberg, 1995) and, for relationships that were significant without the correction, to determine whether the relationship remained significant after applying the correction.

Notes

1. The practice guides offer recommendations on good practices from a panel of experts and review the research evidence on each of those practices using What Works Clearinghouse standards. Practices receive strong ratings when they are supported by studies that meet these standards and show positive impacts and when no studies of similar quality show negative impacts.
2. Wisconsin's RTI Center is a collaboration between Wisconsin's 12 Cooperative Education Service Agencies and the Wisconsin Department of Public Instruction. The center provides technical assistance and guidance to Wisconsin school districts and schools seeking to implement RTI or Positive Behavior Interventions and Supports. More information is available at <http://www.wisconsinrticenter.org>.
3. At the time, school leaders were required to complete the School-wide Implementation Review (the self-assessment mentioned previously), but district leaders were unsure of whether school leaders were completing the review in the same way with commonly held definitions.
4. Positive Pearson coefficients indicate positive relationships (schools with higher values on student or teacher characteristics have higher implementation ratings), and negative coefficients indicate negative relationships (schools with higher values on student or teacher characteristics have lower implementation ratings). More consistent relationships have larger coefficients.
5. For the components of multitiered instruction and evaluation, only the relationships with proficiency rate in math were statistically significant.
6. The district requested that the trainers compress the training from three 8-hour sessions to three 7-hour sessions (to align with the district's mandatory workday). The trainers reported that they were able to cover all of the training topics within the time allotted, with little detriment to quality of the training.
7. The coefficient alpha represents the average of all possible split-half correlations.
8. As noted by Clark and Watson (1995), there is little agreement on the acceptable level of interitem reliability (coefficient alpha). They noted that some psychometricians advocate alphas in the .80 to .90 range as the minimum for acceptability, whereas others have lower standards of .60 or .70. The minimum acceptable interitem reliability used by the What Works Clearinghouse is lower still (alpha = .50). The study team adopted the midpoint between these two standards (.50–.90) as an acceptable level, so as to be in a range that most psychometricians would find acceptable (.70).
9. Partners formerly affiliated with the National Center on Response to Intervention were consulted on these rating labels and agreed on their appropriateness.
10. These cutpoints for classifying implementation fidelity are consistent with the categorization of schools using the National Center on Response to Intervention Integrity Rubric.

References

- Altman, D. (1991). *Practical statistics for medical research*. London, UK: Chapman and Hall.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 289–300.
- Bianco, S. D. (2010). Improving student outcomes: Data-driven instruction and fidelity of implementation in a response to intervention model. *Teaching Exceptional Children Plus*, 6(5), 1–13. <http://eric.ed.gov/?id=EJ907036>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Detgen, A., Yamashita, M., Davis, B., & Wraight, S. (2011). *State policies and procedures on response to intervention in the Midwest Region*. (REL 2011–116). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. <http://eric.ed.gov/?id=ED520326>
- Florida Department of Education. (2014). *Self-Assessment of MTSS Implementation, National pilot version 2.0*. Tallahassee, FL: Author. Retrieved August 20, 2015, from http://www.florida-rti.org/gtips/docs/self_assessment_of_mtss_%28sam%29.pdf.
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, 78(3), 263–279. <http://eric.ed.gov/?id=EJ970680>
- Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, 97(3), 493–513.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.
- Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J., et al. (2009). *Assisting students struggling with mathematics: Response to intervention (RTI) in elementary and middle schools: A practice guide* (NCEE No. 2009–4060). National Center for Education Evaluation and Regional Assistance Working Paper. Washington, DC: U.S. Department of Education, Institute of Education Sciences. <http://eric.ed.gov/?id=ED504995>
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades: A practice guide* (NCEE No. 2009–4045). National Center for Education Evaluation and Regional Assistance

Working Paper. Washington, DC: U.S. Department of Education, Institute of Education Sciences. <http://eric.ed.gov/?id=ED504264>

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness*, 2(1), 88–110. <http://eric.ed.gov/?id=EJ866972>

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, 49(4), 342–352. <http://eric.ed.gov/?id=EJ989963>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

National Center on Response to Intervention. (2010). *Essential components of RTI—A closer look at Response to Intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, and the National Center on Response to Intervention. Retrieved August 20, 2015, from http://www.rti4success.org/sites/default/files/rtiessentialcomponents_042710.pdf.

Rolfhus, E., Gersten, R., Clarke, B., Decker, L., Wilkins, C., & Dimino, J. (2012). *An evaluation of Number Rockets: A Tier-2 intervention for grade 1 students at risk for difficulties in mathematics. Final report*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, Regional Educational Laboratory Southwest. <http://eric.ed.gov/?id=ED529429>

Shah, N. (2011, August 19). Survey of school, district workers shows wider use of RTI. *Education Week: Spotlight on Special Education*. Retrieved July 15, 2016, from http://blogs.edweek.org/edweek/speced/2011/08/yes_another_study_shows_the.html.

Vaughn, S., Denton, C., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading disabilities. *Psychology in the Schools*, 47(5), 432–444. <http://eric.ed.gov/?id=EJ896142>

Viera, A. J., & Garret, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 27(5), 360–363.

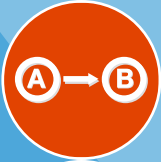
Wisconsin RTI Center. (2015). *School-wide implementation review (SIR), v. 5*. Chippewa Falls, WI: Author. Retrieved February 20, 2016, from <http://www.wisconsinrticenter.org/assets/files/SIR/PrintFriendlySIR.pdf>.

The Regional Educational Laboratory Program produces 7 types of reports



Making Connections

Studies of correlational relationships



Making an Impact

Studies of cause and effect



What's Happening

Descriptions of policies, programs, implementation status, or data trends



What's Known

Summaries of previous research



Stated Briefly

Summaries of research findings for specific audiences



Applied Research Methods

Research methods for educational settings



Tools

Help for planning, gathering, analyzing, or reporting data or research