Learning Disabilities: A Contemporary Journal 10(2), 17-32, 2012

School-wide Student Outcomes of Response to Intervention Frameworks

Daryl F. Mellard
Bruce B. Frey
Kari L. Woods
University of Kansas Center for Research on Learning

Although response to intervention (RTI) has been widely discussed in education literature and is increasingly being implemented in schools throughout the U.S., few studies provide empirical evidence of improved school-wide student outcomes. Presented here are data and an effect size analysis of school-wide student reading gains in five elementary schools implementing RTI components and processes. One school that began the year with below norm skills generally closed the performance gap. Four schools began the year with above norm skills; three of these schools accumulated even greater advantage while one school's advantage diminished. Even as this study offers empirical data on the student outcomes in an RTI setting, more studies are needed before evidential and consequential validity claims can be made about RTI as a school improvement program.

Response to intervention (RTI) was initially conceived as a framework for early intervention among students at risk of reading failure (Deno & Mirkin, 1977; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Fuchs, Fuchs, & Compton, 2004; Torgesen & Davis, 1996). On a concurrent timeline the concept expanded in scope as special educators and others began to see student responsiveness could contribute important information in the identification of specific learning disabilities (SLD) and behavioral disorders (e.g., Fuchs & Fuchs, 1998; Gresham, 2002; Heller, Holtzman, & Messick, 1982; Speece & Case, 2001; Torgesen, et al., 1999; Vaughn, Linan-Thompson, & Hickman, 2003; Vellutino et al., 1996). More recently, claims that RTI is a school improvement program that has the potential to raise all students' achievement have emerged (e.g., Shores & Chester, 2008).

The literature is virtually silent about measures of the school-wide effects of RTI on student achievement. Rather, researchers have primarily concentrated on early intervention to prevent reading or math failures, screening for or predicting at-risk status, and with limited information on its application to SLD identification. Moreover, reported outcomes have been generally limited to such student subpopulations as at-risk students (e.g., Berninger et al., 2006; Bryant et al., 2008; Harn, Linan-Thompson, & Roberts, 2008; Kamps, Abbott, Greenwood, Wills, Veerkamp, & Kaufman, 2008; Schwartz, 2005; Wanzek & Vaughn, 2007, 2008, 2009), students with SLD (e.g., McNamara & Hollinger, 2003; VanDerHeyden, Witt, & Gilbertson, 2007; Vaughn, Linan-Thompson, & Hickman, 2003 Vellutino, Scanlon, Small, & Fanuele, 2006), or English language learners (e.g., Linan-Thompson, Vaughn, Prater & Cirino, 2006; Linan-Thompson, Cirino, & Vaughn, 2007; McMaster, Shu-Hsuan, Insoon, & Cao, 2008; McIntosh, Graves, & Gersten, 2007). Such studies are important elements in validating RTI as a school-wide improvement program, but do not support claims that RTI frameworks raise achievement levels for all students.

Despite the limited research about the broad effect of RTI, the framework has been widely accepted and adopted around the U.S. Local districts and schools have practiced some form of RTI for many years (e.g., Ikeda, Tilly, Stumme, Volmer, & Allison, 1996; Kovaleski & Glew, 2006; Marston, Muyskens, Lau, & Canter, 2002; Mellard, McKnight & Woods, 2009; Telzrow, McNamara, & Hollinger, 2000; VanDer-Heyden et al., 2007). Further, most state education agencies are in some phase of developing or implementing RTI (Berkeley, Bender, Peaster, & Saunders, 2009). This momentum seems to stem not only from organic success in schools and prevalence in the literature, but ultimately from official sanction by IDEA 2004 (P.L. 108-446), which permits RTI as an alternative method of assessing underachievement in SLD identification. Even so, questions about the effectiveness of RTI remain and further study is warranted (Fuchs & Deshler, 2007; Fuchs, Mock, Morgan, & Young, 2003; Grigorenko, 2009).

We assert that the evidential validity of RTI has not yet been established. In 2002, the U.S. Department of Education (USDE) Office of Special Education Programs (OSEP) recognized the importance of having such empirical evidence for RTI, and commissioned the National Research Center on Learning Disabilities (NRCLD) to identify, describe, and evaluate the implementation of RTI in elementary schools. In the context of this broader study, we evaluated school-wide student reading gains for schools where RTI had been sufficiently and verifiably implemented. Thus, while this pre-experimental study will not answer all the questions we have about RTI and student outcomes, the study contributes new data on school-wide effects of RTI on elementary students' reading achievement.

Method

The goal of this study was to measure, analyze, and report school-wide student reading achievement effects of RTI. Because the student outcome data available to us was aggregated by grade level within a school—not student-matched repeated measures—we treated fall and spring data as representing independent groups, despite the fact that most of the students were in both groups. Therefore, we adopted one of Shapiro and Clemens' (2009) suggested indices for evaluating system effects of RTI: rate of improvement across benchmark measures with data aggregated at the school level. The schools selected for this analysis reported data from their progress monitoring processes for this effect size analysis. At the time of the study, RTI was practiced primarily as a reading initiative, thus outcome data is limited to measures of reading skills or subskills.

School selection. Identification of schools actually practicing RTI was perhaps the most important aspect of our study design. We began our study in 2002, when the practice was less prevalent and more ambiguously defined than in the current context. Therefore, we performed a broad search and evaluation process prior to selecting the schools addressed in this study.

In conjunction with the six USDE's federally-funded Regional Resource Centers, we solicited the participation of more than 60 schools in 16 states identified as using RTI components. Forty-one schools responded to the survey (Mellard et al., 2009). A set of open-ended survey and interview items provided the schools' staffs an opportunity to describe and document their RTI design features (e.g., tier structure, screening and progress monitoring instruments and practices, general education curriculum, decision-making rules, special education and learning disabilities determination processes).

Five researchers associated with the NRCLD, each of whom possessed significant research experience and deep familiarity with RTI concepts, acted as jurors and scored the 41 schools' responses. For each school, two of the five jurors completed a scoring rubric that described for each design feature the standards or essentials based on the extant RTI research literature and researcher judgments. The design features on which schools were judged were as follows: (a) school has multiple intervention levels, (b) scientifically-based instruction implemented at each tier, (c) Tier I classroom teachers implement evidence-based instruction/curricula, (d) scientifically-based methods of monitoring students' progress are used, reasonable a priori cut-points or other criteria help determine and distinguish "responsiveness" from "unresponsiveness", (e) scientifically-based instruction implemented (accurately and effectively) in reading, (f) teachers/tutors implement evidence-based instruction/curricula as designed, (g) scientifically-based instruction is implemented with sufficient intensity and frequency to affect change, (h) scientifically-based instruction implemented (accurately and effectively) in reading, (i) progress monitoring occurs in Tier 2, (j) reasonable a priori cut-points or other criteria help determine and distinguish "responsiveness" from "unresponsiveness" in Tier 2, (k) Tier 2 instructional decisions (e.g., return to general education, movement to next level or to special education) are data based and improve learner outcomes, and (l) RTI is used in SLD identification.

Jurors awarded 0 to 2 points for each item in the survey, where responses that earned 2 points provided complete, clear and precise descriptions of replicable RTI procedures, and supplied documentary evidence that verified the practices. The survey and scoring protocols are available from authors on request.

We established two criteria for inclusion in the student achievement effect size phase of the study. These criteria were (a) total scores of 80% or greater on the entire survey from each juror; and (b) an average score of 75% or more on the LD identification feature. Our intention was to ensure that sample schools' procedures were sufficient, even if not complete, to be considered RTI. In three cases with a wide divergence between scores, we had a third juror score the survey, and substituted this score for the outlying score (e.g., juror 1 = 96%, juror 2 = 56%, juror 3 = 48%, so juror 3's score was substituted for juror 1's score). Inter-rater agreement was 83% (n = 23) for the first criteria and 80% (n = 5) for the second criteria. Finally, a panel of jurors performed confirmatory reviews of the five schools that met both criteria and agreed on their selection.

The confirmatory review panel discussed and came to a consensus score for each of the sample schools (Table 1) using a rubric of detailed design features as well as parent and staff interviews. For example, to determine whether the RTI was an effective prevention system, the rubric contained the following standard for school-wide reading assessment:

1. The site obtains reading assessment data or information about reading skill following a designated, fixed schedule.

2. At least 90% of the students participate. Reasons for excluding students from the school-wide screening are reasonable and appropriate, e.g., severe/profound disabilities.
3. Alternative methods to obtain information about reading skill for students excluded from reading assessments have individual curricular relevance and have an index of achievement that allows gains to be measured and evaluated.

Table 1. ***Jurors' Confirmatory and Consensus Score (Percent of Total Points) by School***

	Juror 1	Juror 2	Juror 3	Juror 4	Consensus score
School A	72.5	--	71.7	71.7	--
School B	63.3	--	67.5	--	60.8
School C	85.8	--	85.0	70.0	85.0
School D	73.3	--	--	72.5	74.2
School E	65.8	69.2	--	73.3	--

School settings. School A's 517 students were in grades K–4 with 9% receiving free or reduced lunch. Its RTI model had 4 tiers in which all except 5 students with severe cognitive disabilities were screened for reading risk and placed in appropriate instruction. Seventy-six percent of students received instruction in Tier 1, where *Signatures* reading series (Harcourt Brace) was supplemented by repeated readings, sight word reading, paired reading, and *BlastOff* readers (Bellwether Media). Screening and progress monitoring was based on *Dynamic Indicators of Basic Early Literacy Skills* (*DIBELS*; Good & Kaminski, 2002) oral reading fluency (ORF) benchmarks for 1st and 2nd grades; *DIBELS* ORF benchmarks and percentile ranking (below 56th) on *Terra Nova* (CTB McGraw-Hill) for 3rd grade; and Hasbrouck and Tindal (1992) 50th and 25th percentile rankings in ORF for 4th grade.

School B enrolled 366 students in grades K–5 with 35% receiving free or reduced lunch. Its RTI model had 3 tiers in which all except a few students with low cognitive skills were screened. On average 80% of students received Tier 1 instruction with the following programs and materials: *Literacy Place* (Scholastic), *Read Well K* and *Read Well ®1* (Sopris West Educational Services), *Guided Reading*, *Corrective Reading* and *Corrective Decoding* (McGraw-Hill SRA), *Reading Mastery* (McGraw-Hill SRA), *REWARDS®* (Cambium Learning Group), *Peer-Assisted Learning Strategies©* (Vanderbilt Kennedy Center), and *Read Naturally* (Read Naturally, Inc.). Screening for risk and response to instruction in the core curriculum was determined based on the dual discrepancy data of level and slope (rate of growth) using *DIBELS* measures at all grade levels, and other evidence of responsiveness in reading components.

School C had an enrollment of 977 students located in three facilities with approximately 19% receiving free or reduced lunch. Only data for K-1 grade are reported here. They were implementing a 4-tier RTI model. All students were screened with the exception of a few students with severe disabilities who could not complete the assessment process. Sixty-three percent of students participated solely in Tier 1 core instruction using *Houghton Mifflin Reading* (Houghton Mifflin) and *The Language Tool Kit* (Orton Gillingham), and an additional 23% in *Reading Plus*® (Taylor Associates/Communications Inc.). Screening and response to instruction in the core curriculum was based on level and slope using *DIBELS* benchmarks, and below 70th percentile on local benchmark assessments in K-1; for grades 2-5, CBM fluency levels, Houghton-Mifflin and Gates-McGinitie reading norms.

School D's enrollment of 380 students was in grades K–5 with 91% receiving free or reduced lunch. The school's RTI model had 3 tiers in which all students except those whose parents or IEP team opted for exclusion from the screening process. Core instruction in Tier 1 was provided to 68% of students using *Direct Instruction* (MacMillian-McGraw), *Read Naturally* (Read Naturally, Inc.), *REWARDS*® (Cambium Learning Group), *Open Court Phonemic Awareness and Phonics Kit* (McGraw-Hill), and *Waterford Early Reading Program*™ (Pearson). Screening for risk and response to instruction in the core curriculum was determined by dual discrepancy scores below the 25th percentile on level and the slope on one or more *DIBELS* indices. In addition, the school staff used confirmatory data from CBMs, the *Stanford Achievement Test* (Pearson), *Qualitative Reading Inventory* (Pearson), and *Ekwall Reading Inventory* (Allyn & Bacon) in their decisions.

School E had a total enrollment of 480 students in grades K–5 with 56% receiving free or reduced lunch. Only the 257 K–3 students were included in the school's RTI data. At the time of these activities the school reported only 2 tiers of intervention with 47% of students in Tier 1 and 19% in Tier 2 and the remaining students received instruction through English language learner and special education programs. Core instruction used *Open Court* (McGraw-Hill) curriculum. Students' risk status was based on a dual discrepancy criterion indicated by the level and slope of *DIBELS* benchmarks.

Outcome measures. By design, we took advantage of the RTI screening and/or progress monitoring measurements in place in the five schools. These instruments are listed in Table 2. Many schools used DIBELS, which reports median alternate form reliability by subtest and grade level ranging between .61 and .96; and predictive validity coefficients by subtest and grade level between .17 and .77, as well as a concurrent validity of .90 to .96 among the reading passages used in the 2nd grade oral reading fluency subtest. The DIBELS norming sample was 42% low socio-economic status and 88% racially White. Other schools relied on their own state's assessment systems as the outcome measure, which implies psychometric validity and reliability.

Table 2. *Instruments by school and grade levels*

	Grade levels[a]				
Instrument	School A	School B	School C	School D	School E
DIBELS					
Initial sound fluency		K	K		K
Letter naming fluency		K	K		K
Phoneme segmentation fluency		K, 1	K, 1		K, 1
Nonsense word fluency		K, 1	K, 1		K, 1
Oral reading fluency	1, 2, 4	1, 2, 3, 4, 5			K, 1, 2, 3
Reading achievement tests	3, 4			K, 1, 2, 3, 4, 5	

[a] K= Kindergarten; 1–5 = 1st through 5th grades.

Data analysis. We calculated effect sizes (ES) in quantifying the difference between the one academic year gains of a normative sample and of the students in these schools. That is, we used the tests' normative values to provide an expected rate of improvement. Effect size calculation was the grade level mean difference from the normative sample mean divided by the average standard deviation. An ES value is typically interpreted as .2 = small, .5 = medium and .8 = large. However, we adopted Shapiro and Clemens' (2009) proposed conceptual model for evaluating RTI system effects by comparing rates of improvement for each school to a national normative data set. That is, we compared the effect sizes achieved in one year of instruction in the school's RTI system with the normal growth or effect size for each grade level indicated by each assessment instrument.

Post hoc analysis. To better understand the differences among school outcomes, we also performed a post hoc analysis of the fidelity of implementation scores in (a) general education instruction, (b) tier 2 instruction, (c) special education instruction, and (d) staff preparedness and provision of feedback, which were part of the overall juror scoring criteria. Jurors scored school-reported data and documentation on a 0 to 2 point scale for each of these four domains.

Findings

Table 3. *Kindergarten Mean Score, Standard Deviation and Gain by Instrument and School*

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	*M*	*SD*	N	*M*	*SD*		
DIBELS Initial sound fluency *norms*[b]	*37849*	*12.3*	*10.7*	*38710*	*22.8*	*14.9*	*10.5*	*0.82*
School B	58	16.1	9.8	58	35.8	10.7	19.7	1.92
School C	166	15.7	9.4	170	27.3	13.1	11.6	1.03
School E	58	9.9	6.9	59	21.7	9.4	11.8	1.45
DIBELS Letter naming fluency *norms*[b]	*37396*	*16.0*	*15.3*	*39237*	*44.5*	*18.5*	*28.5*	*1.68*
School B	58	20.9	14.9	58	47.7	12.6	26.8	1.95
School C	167	24.8	16.8	90	49.0	17.4	24.2	1.42
School E	58	12.3	13.1	58	36.0	17.2	23.7	1.56
DIBELS Phoneme segmentation fluency *norms*[b]	*38715*	*27.8*	*20.6*	*39325*	*40.1*	*19.3*	*12.3*	*0.62*
School B	58	29.0	12.7	58	50.9	9	21.9	2.02
School C	170	41.1	14.7	170	54.4	13	13.3	0.96
School E	59	17.3	15.5	58	40.2	14.8	22.9	1.51
DIBELS Nonsense word fluency *norms*[b]	*13221*	*20.1*	*17.8*	*39169*	*32.5*	*22.5*	*12.5*	0.58
School B	58	25.6	16.9	58	38.6	20	13.0	0.70
School C	--	--	--	171	41.9	25	--	--

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]DIBELS norm data are from Good, Wallin, Simmons, Kame'enui, & Kaminski (2002).

Table 4. *First Grade Mean Score, Standard Deviation and Gain by Instrument and School*

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	M	SD	N	M	SD		
DIBELS Phoneme segmentation fluency *norms*[b]	*36865*	*35.2*	*18.9*	*36112*	*50.7*	*14.7*	*15.5*	*0.92*
School B	51	53.0	12.4	51	55.2	8.2	2.2	0.21
School C	129	57.5	10.4	69	63.9	8.6	6.4	0.65
School E	63	32.6	16.4	65	53.1	12.8	20.5	1.41
DIBELS Nonsense word fluency *norms*[b]	*36708*	*30.8*	*22.5*	*36834*	*71.4*	*34.6*	*40.6*	*1.42*
School B	51	43.3	23.1	51	74.8	32.2	31.5	1.14
School C	129	47.5	22.9	94	84.0	33.0	36.5	1.34
School E	63	27.4	29.6	65	78.2	33.5	50.8	1.59
DIBELS CBM Oral reading fluency *norms*[b]	*37410*	*36.9*	*33.1*	*37017*	*60.7*	*38.0*	*33.8*	*0.67*
School A	125	39.0	31.0	125	55.0	33.0	16.0	0.50
School B	51	44.4	30.4	51	57.8	29.6	13.4	0.45
School C	129	57.5	35.6	126	78.1	34.6	20.6	0.59
School E	63	33.1	28.7	65	56.8	27.7	23.7	0.84

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]DIBELS norm data are from Good, Wallin, Simmons, Kame'enui, & Kaminski (2002).

Table 5. *Second Grade Mean Score, Standard Deviation and Gain by Instrument and School*

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	*M*	*SD*	N	*M*	*SD*		
DIBELS CBM Oral reading fluency *norms*[b]	*15494*	*59.6*	*37.5*	*16805*	*100.8*	*42.9*	*41.2*	*1.02*
School A	111	76.0	41.0	111	108.0	44.0	32.0	0.75
School B	67	59.6	26.9	67	87.7	29.2	28.1	1.00
School E	65	51.1	33.4	71	96.9	42.8	45.8	1.20
Reading achievement tests *norms*[c]	*143,033*	*177.2*	*15.0*	*187,912*	*188.2*	*14.4*	*11.0*	*0.75*
School D	70	183.9	15.1	70	196.7	12.2	12.8	0.94

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]DIBELS norm data are from Good, Wallin, Simmons, Kame'enui, & Kaminski (2002). [c]Reading achievement tests norms based on Northwest Evaluation Association (2005).

Table 6. *Third Grade Mean Score, Standard Deviation and Gain by Instrument and School*

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	*M*	*SD*	N	*M*	*SD*		
DIBELS CBM Oral reading fluency *norms*[b]	*10941*	*87.7*	*40.0*	*12531*	*118.5*	*40.4*	*30.8*	*0.77*
School A	131	100.0	32.0	131	123.0	32.0	23.0	0.72
School B	46	98.3	40.2	46	127.5	38.2	29.2	0.74
School E	80	85.0	33.8	84	114.6	32.3	29.6	0.90
Reading achievement tests *norms*[c]	*255,808*	*190.3*	*14.6*	*225,631*	*197.9*	*14.4*	*7.6*	*0.52*
School B	46	192.0	15.1	46	200.0	14.8	8.0	0.54
School D	51	191.0	12.9	51	202.9	9.5	11.9	1.06

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]DIBELS norm based on Good, Wallin, Simmons, Kame'enui, & Kaminski (2002). [c]Reading achievement tests norms based on Northwest Evaluation Association (2005).

Table 7. ***Fourth Grade Mean Score, Standard Deviation and Gain by Instrument and School***

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	*M*	*SD*	N	*M*	*SD*		
Oral reading fluency *norms*[b]	–	*126.0*	–	–	*151.0*	–	*25.0*	–
School A	110	119.0	40.0	110	146.0	43.0	27.0	0.65
School B	70	83.7	31.7	70	114.7	39.9	31.0	0.87
Reading achievement tests *norms*[c]	*250,632*	*199.1*	*14.4*	*226,942*	*205.0*	*14.4*	*5.9*	*0.41*
School B	70	197.0	14.4	70	202.0	12.8	5.0	0.37
School D	72	201.9	14.8	72	209.5	11.3	7.6	0.58

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]Oral reading fluency norms based on Hasbrouck & Tindal (1992) 75th percentile scores. [c]Reading achievement tests norms based on Northwest Evaluation Association (2005).

Table 8. ***Fifth Grade Mean Score, Standard Deviation and Gain by Instrument and School***

	Fall Tests			Spring Tests			Point Gain	Effect Size[a]
Instrument and school	N	*M*	*SD*	N	*M*	*SD*		
Oral reading fluency *norms*[b]	–	*125.0*	–	–	*143.0*	–	*18.0*	–
School B	71	126.9	40.2	71	152.4	46.8	25.5	0.59
Reading achievement tests *norms*[c]	*263,943*	*205.8*	*14.2*	*260,022*	*210.6*	*14.0*	*4.8*	*0.34*
School B	71	209.0	10.0	71	214.0	11.3	5.0	0.47
School D	60	212.2	8.2	60	217.7	8.2	5.5	0.67

Notes: [a]Effect size is mean difference divided by the mean standard deviation. [b]Oral reading fluency norms based on Hasbrouck & Tindal (1992) 75th percentile scores. [c]Reading achievement tests norms based on Northwest Evaluation Association (2005).

School A reported effect sizes that were less than normal annual growth in oral reading fluency (ORF) for 1st, 2nd and 3rd grade (data were insufficient for calculating norm ES for 4th grade ORF). On average 1st grade ORF at the beginning of the academic year was slightly higher than the norm but ended the year slightly lower than norm, which produced an ES of 0.50 compared to a normative ES of 0.67 (Table 4). The 2nd grade students' average ORF at the beginning of year well exceeded norm, but the positive gap closed by year's end, thus the ES was only 0.75 compared to a norm of 1.02 (Table 5). A similar pattern was exhibited by 3rd grade students whose fall scores well exceeded norm and spring scores were slightly higher than norm; the ES was 0.72, just slightly less than norm of 0.77 (Table 6). Lastly, 4th grade students were the only group in this school to demonstrate greater than normal growth (Table 7). They began the year reading slightly below a normal ORF rate, and ended the year slightly above the norm for an ES of 0.65. DIBELS norm data are not available for 4th or 5th grade ORF, thus norm ES could not be calculated. For comparison purposes we present Hasbrouck and Tindal's (1992) 75th percentile oral reading fluency rates.

School B demonstrated larger than expected ES for kindergarten and 5th grade student achievement, a below norm ES on multiple measures in 1st grade, and approximately the same as norm ES in 2nd, 3rd and 4th grades. Kindergarten initial sound fluency (ES = 1.92 vs. 0.82), letter naming fluency (ES = 1.95 vs. 1.68), phoneme segmentation fluency (ES = 2.02 vs. 0.62) and nonsense word fluency (ES = 0.70 vs. 0.58) growth seem to indicate that the tiered instruction was beneficial to the kindergarten class as a whole (Table 4). The below norm effect sizes in 1st grade seem to be a product of ceiling effects. For example, students began the year on average with phoneme segmentation scores that were higher than year-end norms, and thus had an ES = 0.21 compared to norm of 0.92. With nonsense words and ORF measures, students also began the year with higher average scores than norms and ended the year slightly above norms, resulting in ES also slightly below norms (phoneme segmentation fluency ES = 1.14 vs. 1.42 and ORF ES = 0.50 vs. 0.67). Likewise, 5th grade reading achievement test scores showed better than normal gains (ES = 0.47 vs. 0.34).

School C reported outcomes for only kindergarten and 1st grade students. Kindergarten students began the year with higher than norm initial sound fluency, letter naming fluency, and phoneme segmentation average scores and demonstrated more than expected gains on two of the three skills (initial sound fluency ES = 1.03 vs. 0.82 and phoneme segmentation fluency ES = 0.96 vs. 0.62). Letter naming fluency growth was sufficient to remain above normal, but may have reached a ceiling for a slightly smaller than norm ES (ES = 1.42 vs. 1.68). The 1st grade average scores on each measure began and ended the year higher than norm, and each ES was slightly below norm as a result (phoneme segmentation fluency ES = 0.65 vs. 0.92; nonsense word fluency ES = 1.34 vs. 1.42; and ORF ES = 0.59 vs. 0.67).

School D did not report outcomes for kindergarten or 1st grade students. However, the school's 2nd through 5th grade students performed above norm on beginning and end of year reading achievement tests. Further, RTI instruction resulted in rates of growth with ES greater than norm at every level (2nd grade ES = 0.94 vs. 0.75; 3rd grade ES = 1.06 vs. 0.52; 4th grade ES = 0.58 vs. 0.41; and 5th grade ES = 0.67 vs. 0.34).

School E students at every grade level began the school year with below average scores and ended the year with scores approaching or exceeding the norm. Two of the three kindergarten ES were far greater than the norm (initial sound fluency ES = 1.45 vs. 0.82 and phoneme segmentation fluency ES = 1.51 vs. 0.62), and the third measure, letter naming fluency, was near the norm level (ES = 1.56 vs. 1.68). First grade students improved their average scores in reading subskills to above norm levels with larger than normative ES (phoneme segmentation fluency ES = 1.41 vs. 0.92 and nonsense word fluency ES = 1.59 vs. 1.42). Oral reading fluency average scores, while not ending above norm levels, grew more than the norm (ES = 0.84 vs. 0.67). The 2nd grade and 3rd grade students followed a similar pattern of achievement on measures of ORF increasing to near norm levels (2nd grade ES = 1.20 vs. 1.02 and 3rd grade ES = 0.90 vs. 0.77).

Fidelity of implementation. For many educational innovations a tension frequently exists between maintaining fidelity of implementation and adapting components of the intervention to fit an educational context (Hulleman & Cordray, 2009). Indeed, even after all the reviews and judgments about the schools in this study, the best of the five schools described here had a 72% fidelity rate and the other four schools were rated between 39% and 49%. Given that these schools were relatively early adopters of RTI, we are not surprised at these fidelity scores. We present these ratings here not to cast doubt on the schools' RTI programs, but to lend insight into some of the differences among school outcomes (Table 9). For example, School A, the only school that did not demonstrate better than expected gains in student reading, scored only 0.5 out of a possible 2 points on general education core instruction.

Discussion

Three types of results emerge from the data that describe student reading achievement or skills in these five schools that were judged to sufficiently and verifiably implement RTI. The first type of result was accumulated advantage, analogous to the Matthew Effect (Stanovich, 1986), wherein the rich get richer. Schools B, C, and D students generally scored above the norm on their fall tests and not only maintained this advantage but gained more than expected during the year of tiered instruction. These schools might be characterized as high achievers and high gainers. Although one can expect some regression to the mean on subsequent measures, this effect was not noted, which suggests that noteworthy instructional programming was effectively supporting their students. The second kind of outcome is demonstrated by School E, in which students began the year averaging well below normal and made substantial gains that closed the gap between them and the test's normative sample, and in some instances exceeded norm. School E can be characterized as the lower performing but higher gaining school. A third, less positive result, is evident in School A data, where the average fall test scores were often above the norm but students did not maintain this advantage in spring test scores. School A is disconcerting. We suggest that in this school, the staff should carefully review other school-wide and grade level assessment data to determine the consistency of this decreasing achievement. Further, the staff might examine the fidelity with which the evidence-based instructional practices and curricular materials are delivered. Although our study lacked the data to exam the possibility, one might argue that Schools A and E are examples of regression to the mean. For School E spring score increases moved them closer to the mean, and for School A spring score decreases moved them closer to the mean.

We speculate that School A's general education core instruction was in some way inferior to the other four schools based on the juror's lower ratings on fidelity of implementation in general education instruction (see Table 9). This finding highlights the importance of sound core reading instruction, which is intended to support the majority of the students, within an RTI framework.

Table 9. *Average fidelity of implementation scores by domain*

	School A	School B	School C	School D	School E
General education instruction	0.50	1.50	1.00	0.75	1.17
Tier 2 instruction	1.00	1.33	0.67	0.75	1.33
Special education instruction	0.67	1.50	0.67	0.75	0.50
Staff preparedness and provision of feedback	1.00	1.50	1.33	1.25	0.83
Total fidelity score	3.17	5.83	3.67	3.50	3.83
Fidelity percentage (school score ÷ maximum score)	39%	72%	46%	44%	49%

Conclusion

Educators and the public at large face increasing challenges as they work to meet the needs and improve the academic and behavioral outcomes of an increasingly diverse student population and to do so with greater demands of accountability and reduced resources. Some educators consider the RTI framework as offering a significant opportunity for schools to meet the needs of all their students. The RTI framework might accomplish these outcomes through its emphasis of responding to data on student responsiveness to the schools' and teachers' curricular and instructional decisions. Considering RTI as a prevention framework, students who are at-risk for academic and behavioral difficulties (e.g., truancy, suspension, and dropping out) are identified through the students' scores on periodic screenings and progress monitoring. In these schools' implementation, the students' reading scores across multiple grades indicated some improvements. We also noted a school in which the improvements were not maintained and possibly linked to an inadequate core curriculum or its delivery. As a pre-experimental design, the results are suggestive but could have been influenced in multiple ways by numerous threats to their validity. Perhaps the bottom line is that these schools' varied RTI frameworks and situations are supporting students and leading to positive outcomes. We await further, more rigorous research designs to more fully evaluate RTI's evidential and consequential validity.

References

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

Berninger, W., Rutberg, J., Abbott, R., Garcia, N., Anderson-Youngstrom, M., Brooks, A., & Fulton, C. (2006). Tier 1 and Tier 2 early intervention for handwriting and composing. *Journal of School Psychology*, *44*(1), 3–30.

Bryant, D., Bryan, B., Gersten, R., Scammacca, N., Funk, C., Winter, A., Shih, M., & Pool, C. (2008). The Effects of Tier 2 Intervention on the Mathematics Performance of First-Grade Students Who Are at Risk for Mathematics Difficulties. *Learning Disability Quarterly*, *31*(2), 47–36.

Center for Teaching and Learning, University of Oregon. (2002). *Assessment committee analysis of reading assessment measures: Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* 5th Edition. Eugene, OR: Author. Available at https://dibels.uoregon.edu/techreports/dibels_5th_ed.pdf

Coyne, M., Kaméenui, E., Simmons, D., Harn, B. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention, *Journal of Learning Disabilities*, *37*(2), 90–104.

Deno, S., & Mirkin, P. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Foorman, B., Francis, D., Fletcher, J., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children, *Journal of Educational Psychology*, *90*(1), 37–55.

Fuchs, D., & Deshler, D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*(2), 129–136.

Fuchs, L.S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13*, 204–219.

Fuchs, D., Fuchs, L. & Compton, D. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disabilities Quarterly*, 27(4), 216–227.

Fuchs, D., Mock, D., Morgan, P., & Young, C. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, *18*(3), 157–171.

Good, R. & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved June 7, 2006 from http://dibels.uoregon.edu/measures/orf.php

Good, R., Wallin, J., Simmons, D., Kame'enui, E., Kaminski, R. (2002). *System-wide percentile ranks for DIBELS benchmark assessment* (Technical Report 9). Eugene, OR: University of Oregon.

Gresham, F.M. (2002) Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.), *Identification of Learning Disabilities.* (pp. 467 – 519). Mahwah, NJ: Lawrence Erlbaum Associates.

Grigorenko, E. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, *42*(2), 111–132.

Harn, B., Linan-Thompson, S., & Roberts, G. (2008). Intensifying instruction: Does additional instructional time make a difference for the most at-risk first graders? *Journal of Learning Disabilities*, *41*(2), 115–125.

Hartman, W. T., & Fay, T. A. (1996). Cost-effectiveness of instructional support teams in Pennsylvania. *Journal of Educational Finance, 21,* 555–580.

Hasbrouck, J., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children*, *24*(3), 41–44.

Heffner, C. (2004). *Research Methods*. Website: AIIPsychONLINE: The virtual psychology classroom. http://allpsych.com/researchmethods/preexperimentaldesign.html

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

Hulleman, C., & Cordray, D. (2009). Moving from the lab to the field: The Role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88-110.

Ikeda, M., Tilly, W.D., Stumme, J., Volmer, L., & Allison, R. (1996). Agency-wide implementation of problem solving consultation: Foundations, current implementation, and future directions. *School Psychology Quarterly*, *11*(3), 228–243.

Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences for students most at risk in kindergarten: Two-year results for secondary- and tertiary-level interventions. *Journal of Learning Disabilities*, *41*(2), 101–114.

Kovaleski, J., Gickling, E. E., Morrow, H., & Swank, H. (1999). High versus low implementation of instructional support teams. *Remedial and Special Education, 20*(3), 170–183.

Kovaleski, J., & Glew, M. (2006). Bringing instructional support teams to scale: Implications of the Pennsylvania experience. *Remedial and Special Education*, *27*(1), 16–25.

Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities*, *39*(5), 390–398.

Linan-Thompson S., Cirino, P., & Vaughn, S. (2007). Determining English language learners' response to intervention: Questions and some answers. *Learning Disabilities Quarterly*, *30*(3), 185–195.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2002). Problem solving model for decision-making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice*, *18*(3), 187–200.

McIntosh, A., Graves, A., & Gersten, R. (2007). The effects of response to intervention on literacy development in multiple-language settings. *Learning Disability Quarterly*, *30*(3), 197–212.

McMaster, K., Shu-Hsuan, K., Insoon, H., & Cao, M. (2008). Peer-assisted learning strategies: A "Tier 1 "approach to promoting English learners' response to intervention. *Exceptional Children*, *74*(2), 194–214.

McNamara, K., & Hollinger, C. (2003). Intervention-based assessment: Evaluation rates and eligibility findings. *Exceptional Children*, *69*(2), 181-193.

Mellard, D.F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice*, *24*(4), 186–195.

Northwest Evaluation Association. (2005). *RIT scale norms: for use with measures of academic progress and achievement level tests*. Lake Oswego, OR: Author.

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade, *Learning Disability Research & Practice*, *15*(1), 43–54.

Shapiro, E., & Clemens, N. (2009). A conceptual model for evaluating system effects of response to intervention. *Assessment for Effective Intervention*, *35*(1), 3–16.

Schwartz, R. (2005). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, *97*(2), 257–267.

Shores, C., & Chester, K. (2008). *Using RTI for school improvement: Raising every student's achievement*. Thousand Oaks, CA: Corwin Press.

Speece D., & Case, L. (2001). Classification in context: An alternative approach to identifying early reading disability, *Journal of Educational Psychology*, *93*(4), 735–749.

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

Telzrow, K., McNamara, C., & Hollinger, C. (2000). Fidelity of problem-solving implementation and relationship to student performance. *School Psychology Review*, *29*(3), 443–461.

Torgesen, J. & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness, *Journal of Experimental Child Psychology*, *63*(1), 1–21.

Torgesen, J., Wagner, R., Rashotte, C., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction, *Journal of Educational Psychology*, *91*(4), 579–593.

VanDerHeyden, A., Witt, J., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a Response to Intervention (RTI) model on identification of children for special education. *Journal of School Psychology*, *45*(2), 225–256.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities, *Exceptional Children*, *69*(4), 391–409.

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

Vellutino, F., Scanlon, D., Small, S., & Fanuele, D. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities. *Journal of Learning Disabilities*, *39*(2), 157–169.

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

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, *41*(2), 126–142.

Wanzek, J., & Vaughn, S. (2009). Students demonstrating persistent low response to reading intervention: Three case studies. *Learning Disabilities Research & Practice*, *24*(3), 151–163.

Acknowledgment

The primary data reported in this manuscript were collected in a study funded by U.S. Department of Education, Office of Special Education Programs (Grant No. H324U010004).