

**Much Ado About Little: The Dangers of Disseminating the RTI  
Outcome Study without Careful Analysis**

Mark R. Shinn, Ph.D.

National Louis University

Rachel Brown, Ph.D.

University of Southern Maine

Corresponding author:

Mark R. Shinn

National Louis University

[mshinn@nl.edu](mailto:mshinn@nl.edu)

847.275.7200

## **Much Ado About Little: The Dangers of Disseminating the RTI Outcome Study without Careful Analysis**

### **Introduction**

In November 2015, the US Department of Education's Institute for Education Sciences (IES) released its report on the effectiveness of a significant change in educational practices entitled *Evaluation of Response to Intervention Practices for Elementary School* (Balu, Zhu, Doolittle, Schiller, Jenkins, & Gersten, 2015). Almost immediately, a number of influential education dissemination periodicals released their interpretation of the IES Report's results. For example, *Education Week*, a digital publication that attracts more than 1.5 million readers per month, published a story with the headline "*RTI Practice Falls Short of Promise: First Graders Who Were Identified for More Help Fell Further Behind*" (Sparks 2015). The first two paragraphs of the narrative stated

"the most comprehensive federal evaluation of the approach to date finds that it may hold back some of the children it was originally designed to support. First graders who received reading interventions actually did worse than virtually identical peers who did not get the more targeted assistance."

Similarly, another article in *EdCentral*, with the title "*New study raises questions about RTI implementation*" (Loewenberg 2015) reported that

"The second and third grade students who received Tier 2 supports experienced no significant reading benefits from the interventions...first graders who received Tier

2 interventions actually *did worse* (their emphasis) than their peers who did not receive targeted assistance.”

In a number of listservs and discussions, the articles reporting the IES study were widely disseminated with further editorializing about RTI’s “failure” and that it may harm students.

### **Purpose of the Paper**

This article’s purpose is not to attack either the original study or some of the interpretations of the study’s results. We appreciate, and admire, any effort to investigate the school-based implementation of a major reform such as RTI on such a large scale in which the investigators were constrained by a retrospective analysis with no control over implementation. Evaluating RTI practices of a pool of 146 schools distributed across 13 states with significant differences in their measures and methods, requires significant *compromises* to draw any conclusions. The most obvious compromises were made in (a) the selection of the students whose achievement was investigated (i.e., students at the grade-level achievement margins), and (b) the operationalization of RTI through use of surveys and self-reports rather than of direct observations. Furthermore, for any follow-up reports or for readers, it is a formidable task to try to accurately report the results of such a large scale study that is described in a 308-page document with a 17-page executive summary, a 121-page narrative with 61 tables, and 31 figures.

This paper is written to urge educators to engage in a comprehensive critical analysis of the full report, including a number of important cautions and concerns expressed by the report’s authors, themselves. We encourage careful attention to two major concerns we have with the paper and subsequent outcome reporting so that an error so common to education, abandoning important reforms, is not repeated again. The compromises required to complete this study

should not comprise schools' efforts to improve outcomes for students, especially those students with significant achievement discrepancies.

### **The Devil is in the Details**

A technical and detailed critique of the report could focus on any number of threats to internal validity (Shadish, Cook et al. 2002) concerns. Among these concerns one could consider instrumentation where subjects were identified by the outcomes of the different classification schemes used by different schools with different tests and with unspecified criteria.

Other instrumentation concerns include the following:

- Outcomes were judged by a single post-test measure at Grade 2, two post-test measures at Grade 1, and the use of 13 different state standards tests as post-tests at Grade 3.
- All of the intervention variables described in the report were based solely on survey reports from the sites with no observational data to cross validate the self-report data.
- The complete lack of any fidelity of RTI implementation data.

Instead, in this paper we focus solely on what we consider the two foremost concerns where we question:

1. Where these the “right” subjects with which to judge RTI outcomes (External Validity);  
and
2. Were the interventions delivered by the schools and evaluated by the report what most educators would consider “RTI? (Internal Validity and Inadequate Preoperational Explication of Constructs or Construct Validity of Putative Cause).

**External Validity: Studying RTI Impact on the “Right” Students**

Given that Balu et al. (2015) study was conducted well after the participating schools’ RTI implementation began, it was not possible to use an experimental design with random assignment. The lack of any common pre-test measure(s) among other implementation issues precluded a variety of quasi-experimental designs that might allow defensible conclusions about cause and effect. One option in these circumstances was the regression discontinuity (RD) design that formed the basis for analyzing the achievement outcomes. RD provides a way to compare effects between groups when a true experimental design is not possible (Imbens & Lemieux, 2008). It does this by using a preset cut score to define the *threshold* of participants who do or do not meet a specific criterion, in this instance, those students whose screening scores led to RTI intervention. In relation to the RTI practices evaluated in the Balu et al. study, the RD design was set up such that the threshold was the cut point between students who were at the screening margins for *grade-level or below-grade-level reading achievement*.

Most, but not all, of the students with scores just below the screening cut point (i.e., Somewhat Below) were reported to receive Tier 2 intervention, and most of the students with scores just above the cut score (At or Somewhat Above) received Tier 1 (core instruction). Students with severe reading difficulties (Far Below) *were excluded* from the evaluation of RTI effects.

Differences between the two groups’ achievement relations from screening scores to post-test scores allow statements of cause and effect. It will be noted later in this paper that the conclusions about achievement differences was confounded because the distinction between groups was not one of dichotomized RTI intervention or no RTI intervention. The strongest conclusions about cause and effect in RD could be made when students:

1. Above the margins received Tier 1 *only* instruction, and
2. Below the margins received additional RTI intervention.

Unfortunately, large number of students above the margins, up to 45% also received additional intervention as part of RTI.

This intervention confound aside, the use of a RD design was appropriate, given the constraints that the investigators faced. However, the design required a *compromise* with respect to the target population of students who may require more intensive RTI intervention. From its earliest implementation, (e.g., Germann & Tindal 1985; Marston & Magnusson, 1985), the target population for screening and early intervention has been students *significantly discrepant* in their achievement from typically developing peers, not students at the margin of grade-level achievement. RTI historically has been designed to increase the range of general education options for significantly discrepant students, the students who might meet the study's criterion for Far-Below grade level achievement margins, rather than requiring schools to use special education as the only service delivery option. Only in the last decade have serious attempts been made to expand the options for early intervention for less discrepant students.

The investigators *could* have compared achievement outcomes with students *Far Below the grade-level achievement margin*, but this choice would have required them to identify students above that margin who *did not* receive RTI intervention. Thus, the design required a significant compromise in terms of the students investigated.

We maintain that drawing conclusions about RTI effectiveness based on the students at the margins of grade-level achievement requires great caution. Balu et al. were clear about overgeneralizing the results to *all* students who may receive RTI intervention, stating that the

“design determines that the impact findings are applicable *not to everyone* (our emphasis) receiving either Tier 2 or Tier 3 intervention, but *only to students whose fall screening scores were close to the cut point*” (our emphasis) (ES-p. 6).

### **Internal Validity and Construct Validity of Putative Cause: Was Intervention “RTI?”**

One critical component of internal validity is confidence in *the construct validity of putative cause* or the potential for *inadequate preoperational explication of constructs* (Shadish, Cook, & Campbell, 2002). That is, to conclude “RTI” (i.e., the independent variable) failed to impact the achievement of RTI students, one must have confidence that what was delivered was “RTI.”

The Report averages a number of RTI intervention features across 126 schools and we believe the Report best provides compelling evidence of *average weak RTI implementation* falling well short of the critical features of quality RTI described in a substantial number of professional resources (e.g., Burns & Gibbons, 2012; Fuchs & Fuchs, 2007; Jimerson, Burns, & VanDerHeyden, 2007) and prominent websites, such as the *National Center for Response to Intervention* ([www.RTI4success.org](http://www.RTI4success.org)), the *National Center for Intensive Intervention* ([www.intensiveintervention.org](http://www.intensiveintervention.org)) and the *RTI Action Network* ([www.RTInetwork.org](http://www.RTInetwork.org)) describe the common features of RTI. The shortfalls against these expected features of quality RTI in the Report makes us question whether it is legitimate to draw inferences about its effects.

### **Intervention Intensity Matches Student Need**

The Report’s descriptive data on RTI intervention characteristics made it impossible to determine any differences in intensity based on severity of students’ reading discrepancies. Whereas the evaluation of RTI’s effect on achievement are limited to Somewhat Below grade-level students, the RTI intervention descriptions are based on *all students who received RTI*. As the Report states, “groups serving readers Somewhat Below and Far Below grade level are

*combined* (our emphasis) to represent those students who are most likely to be in the “treatment” group... and then are compared with groups serving readers At or Above grade level” (p. 58).

By combining the data on RTI interventions of students Somewhat Below *and* Far Below grade-level margins, the practices described are likely over-estimates of intensity. If RTI were implemented with fidelity, then the interventions for students Far Below grade-level achievement margins would be much more intense than the interventions for Somewhat Below the margin students. To us, it is plausible that in actuality, intervention intensity for the students Somewhat Below students that they actually received are *even less* than the intensity described in the Report.

### **RTI Intervention That Is Different**

The data in the Report provides evidence that there was *little difference* between the reading instruction that students Above (no RTI Intervention) and Somewhat Below the grade-level margins (RTI intervention). As a point of fact, the Report details that students in *both groups* received additional intervention through their RTI implementation model. At the minimum, one-third to almost one half of the schools reported that they provided additional intervention to *all* students beyond core instruction. In other words, up to 50% of students who scored above the criterion for grade-level proficiency also received additional intervention. The confound of “additional intervention” to both comparison subjects presents two problems. First, it makes it difficult to parse out differences from “additional intervention” from “RTI intervention.” Second, providing additional intervention to all students has the potential to dilute intervention intensity and resources if distributed to all students rather than a portion of students. Importantly, it should be noted that the grade where additional intervention was delivered to all

students was highest was Grade 1, which is where the Report identified negative achievement outcomes on one of the measures.

The details about intervention practices in Chapter 4 also raise concerns about when schools provided additional intervention that much of it was above and beyond Tier 1 core instruction. As shown in Table 4.6 of the report (Balu et al., 2015, p. 66),

"[i]n schools that provided intervention only to readers below grade level in the corresponding grade, those groups received 89 minutes per week of small-group instruction time, compared with 62 minutes for groups at or above grade level" (Balu et al., 2015, p. 66).

The difference in time between groups equals 5.4 more minutes per day for first-grade students Somewhat *and* Far Below the grade-level margins. In the schools that provided intervention to all students, first-grade small group instruction differed by 8 additional minutes per day. Compared with prior early intervention research, it is unlikely that the addition of so few minutes of additional small-group instruction would produce significant differences in reading gains, especially for students with severe reading difficulties.

### **Intervention is Typically Supplemental**

The Report provided considerable evidence that RTI intervention *supplanted* core instruction, a contraindication of high quality implementation where Tier 2 is typically *supplemental* to core instruction. For *most* students who received RTI, at least part of their RTI intervention *supplanted* rather than supplemented core instruction, a problem noted by the Report's authors. According to the report, "[t]he implication is that intervention may have replaced rather than supplemented some instruction services during the core" (p. 60). The

percentages of supplanted instruction approximated 67% of students across grades, which once again included both the Somewhat Below and Well Below intervention groups.

Among the schools that provided intervention *only* to those students Somewhat Below the grade-level margins, an additional 167 to 187 minutes per week was reported, with the lowest amount of intervention time at Grade 1. The additional intervention time for first graders translates into an average of an additional 33 minutes per day. However, it was unclear in the tables if this intervention time supplanted or was supplemental instruction to the core. This problem was noted by the Report authors, when they stated that “some intervention groups occurred during the core and may have displaced some instruction time, rather than supplementing it” (p. 69). It is important to once again note that these additional intervention time figures are (a) from the intervention-only schools that comprised about half of the sample, and (b) the intervention time reported was based on data that included students Far Below grade level.

### **Interventions are Delivered by Highly-Trained Personnel**

Another RTI characteristic is that specialized instruction is delivered by highly-trained personnel. The Report suggests that this characteristic was the exception rather than the norm. Analysis of the data reported in Appendix Table C.9 of the Report suggests that for the all-student intervention schools, RTI interventions were provided mostly by either the classroom teacher or a paraprofessional 64% to 77% of the time. In the schools that provided intervention only to Somewhat Below-margin students, this percentage decreased from 61% to 64%. Regardless of RTI model, interventions delivered by classroom teachers constituted the greatest single intervention provider, but paraprofessionals, not highly trained personnel, were the second most common. All-school and Below-Level Only intervention schools ranged from 20% to 31%

at the former to 22% to 37% at the later. Regardless of RTI approach, the highest proportion of students receiving instruction by paraprofessionals was at Grade 1 where the Report found potential negative RTI impact. If reading specialists and special education teachers are examples of highly trained personnel, at best, 23% of intervention was provided by these school personnel (Grade 2 below-only schools) to as little as 9% (Grade 3 below-only).

### **Intervention Effects Are Closely Monitored**

Frequent progress monitoring practices using scientifically sound tests not only is a feature of quality RTI implementation, but also effective educational practices (Hattie 2009). The Report provides some strong evidence that nearly all schools engaged in frequent progress monitoring practices using measures of oral reading fluency, which is a scientifically sound practice reviewed favorably in the RTI literature. However, it also reported a very high frequency of scientifically unsound practices such as use of running records, ranging from 70% to 78% of schools in the below-grade level intervention only schools and even higher in the all-students intervention schools (83%-89%).

### **Conclusions and Concerns**

The Report's executive summary (Balu et al, 2015, p. 1), stated that

This report provides new information on the prevalence of RTI *practices* in elementary schools, illustrates the implementation of RTI *practices* for groups of students at different reading levels, and provides evidence on effects of one key element of RTI: assigning students to receive reading intervention services (emphasis added) (Balu et al, 2015, p. 1),

Therefore, the Report's primary focus was on RTI *practices*, which makes sense in relation to the nature of the methods used and data available to the investigators. Nonetheless, media reports

about the report have focused primarily on the reported lack of or negative student reading outcomes attributable to RTI. Yes, students' changes in reading skills were a component of the Report. However, concerns about external validity and internal validity must temper the interpretation of outcomes. We argue that this report is best understood as a study of RTI implementation practices and should be interpreted from an implementation quality or fidelity perspective. As noted in chapter 3 of the Report, schools considered the interventions they were delivering to students as RTI, stating "[t]he majority of reference sample schools adopted an RTI framework, including implementing multiple reading tiers and using data-based decision-making practices" (Balu et al, 2015, p. 27). However, these features alone are only the broad brush stroke of RTI and the average intensity of intervention features described in this Report are disappointing. On whole, they clearly suggest that in in many schools, there is a long way to go for high quality RTI.

Given the subjects whose achievement gains were the basis for the conclusions, students Somewhat Below grade-level margins, we worry that educators who fail to read the full report carefully will reach the conclusion that RTI doesn't work or will make students worse. The Report's authors are clear about the groups to which the RTI outcomes should be generalized. As the study's authors note,

“This [study's] design determines that the impact findings are applicable *not to everyone receiving either Tier 2 or Tier 3 intervention, but only to students whose fall screening scores were close to the cut point*. Students close to the cut point are largely Tier 2 students but also include a small portion of Tier 3 students” (Balu, et al., 2015, p. ES-6; emphasis added).

Sadly, secondary reports on this Report have failed to emphasize this statement from the

Executive Summary. Given that RTI typically is intended *not* for students at the margins of *average* performance, but students with more significant reading achievement deficits, external validity must be considered when interpreting and disseminating the findings. Furthermore, given the descriptions of the differences in intervention intensity between these students, one must question whether the intervention(s) provided to these students at the margin would be considered “RTI.”

### **Precedent of Over-Generalization of Previous Impact Reports**

The challenges facing quality implementation of RTI closely parallel another major national educational improvement effort, *Reading First*. Like RTI, *Reading First* challenged the conventional practices of reading instruction. Instead of a whole language approach, where students were taught by experience and exposure with authentic text, *Reading First* required more carefully designed, research-based core curricula and explicit, teacher-led instruction. Like RTI, *Reading First* challenged how educators *thought* about reading instruction and how students learn. Specifically, these efforts challenged the idea that learning to read is a natural process much like learning to talk, and that phonics “deadens them to the joys of reading” (Finn and Petrilli, 2008). Due to the disconnect between most teachers' expectations about reading instruction, Finn and Petrilli noted that they “watched in horror as *Reading First* was attacked, defanged, and eventually brought to its knees” (p. 5).

One way that the *Reading First* initiative was ultimately terminated was the widespread media reporting that it did not work. For example, *EdWeek* reported the results of the *Reading First Impact Study: Interim Report* (Gamse, Bloom, Kemple, Jacob) in April 2008 as “[t]he \$1 billion-a-year Reading First program has had no measurable effect on students’ reading comprehension, on average” (Kennedy Manzo 2008). Nonetheless, the final *Reading First*

*Impact Study* from November 2008 released months later showed there were numerous benefits from *Reading First* (Gamse, Jacob, Horst, Boulay, & Unlu, 2008). How the findings of major educational research are reported in both mainstream and specialty media can affect educators' perceptions of the outcomes. This appears to have been the case for *Reading First* as well as for the recent RTI study.

### **Challenges of Efficacy and Effectiveness**

The effects of early and powerful intervention on the reading achievement of significantly discrepant students have a lengthy and positive history. In carefully designed studies, (e.g., Denton, Fletcher, Anthony, & Francis, 2006; Torgesen, 1998; 2001; 2002; 2004a; 2004b) student reading achievement improved significantly for students with severe reading discrepancies versus non-treated students. Of course, this observed *efficacy* -- how well an intervention actually works under idealized conditions such as carefully controlled research studies -- must be compared to *effectiveness* which is how well it works under normal, routine conditions of usage (Walker & Shinn 2010). Given the reported characteristics of the interventions delivered to students at the margins in this study, it appears that there were intervention features that many of the impact schools had difficulty putting into practice. This problem becomes one of attending to implementation science, a solution that may be short-circuited if the IES study results are not interpreted and disseminated carefully.

### References

- Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of Response to Intervention Practices for Elementary School Reading* (NCEE 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovalski, J., Prasse, D., et. al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.
- Burns, M.K., & Gibbons, K. (2012). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices (2nd ed.; School-Based Practice in Action)*. New York: Routledge.
- Denton, C.A., Fletcher, J.M., Anthony, J.L., & Francis, D.J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities, 39*, 447–466.
- Finn, C. E. & Petrilli, M.J. (2008). *Forward. Too good to last: The true story of Reading First*. S. Stern. Washington DC, Thomas B. Fordham Institute: 1-8.
- Fuchs, L.S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *TEACHING Exceptional Children, 39(5)*, 14-20.
- Gamse, B. C., Bloom, H. S., Kemple, J. J., & Jacob, R. T. (2008). *Reading First Impact Study: Interim Report* (NCEE 2008-4016). Washington, DC: Institute of Education Science.
- Gamse, B. C., Jacob, R. T., Horst, D. P., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study: Final Report* (NCEE 2009-4038). Washington, DC: Institute of Education Science.

Germann, G. & Tindal, G.A. (1985). "Applications of direct and repeated measurement using curriculum based assessment." *Exceptional Children* 51(2): 110-121.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY, Routledge.

Jimerson, S.R., Burns, M.K., & VanDerHeyden, A.M. (2007). *Handbook of response to intervention: The science and practice of assessment and intervention*. New York: Springer.

Kennedy Manzo, K. (2008). "Reading First Doesn't Help Pupils 'Get It'." from <http://www.edcentral.org/RTIstudy/>.

Loewenberg, A. (2015). "New study raises questions about RTI implementation." from <http://www.edcentral.org/RTIstudy/>.

Marston, D. and D. Magnusson (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children* 52: 266-276.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sparks, S. D. (2015). RTI Practice Falls Short of Promise, *Education Week*.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator*, Spring, 32-39.

Torgesen, J. K. (2001). The theory and practice of intervention: Comparing outcomes from prevention and remediation studies. In A. J. Fawcett & R. I. Nicolson (Eds.), *Dyslexia: Theory and good practice* (pp. 185-201). London: David Fulton Publishers.

Torgesen, J. K. (2002). Lessons learned from intervention research in reading: A way to go before we rest. *Learning and Teaching Reading*, 89-103.

Torgesen, J. K. (2004a). Lessons learned from the last 20years of research in the interventions for students who experience difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 225-229). Baltimore, MD: Brookes Publishing.

Torgesen, J. K. (2004b). *Teaching every child to read: The knowns and the unknowns*. National Association of School Psychologists. Dallas, TX.

Walker, H. M. & Shinn, M.R. (2010). Systemic, evidence-based approaches for promoting positive student outcomes within an RTI framework: Moving from efficacy to effectiveness. In M. R. Shinn and H. M. Walker (Eds.) *Interventions for achievement and behavior problems in a three-tier model, including RTI*. Bethesda, MD, National Association of School Psychologists: 1-26.

*Disclosures*. Mark R. Shinn, Ph.D. serves as a consultant to Pearson Assessment on their *aimsweb* Curriculum-Based Measurement (CBM) progress and screening products and information management systems.