Promise, Practicality and Prudence: A Synthesis of Second Generation RTI Research

Howard Weiner

I Importance of RTI Research and impact on Student Outcomes

The primary purpose of research and assessment in schools should be to help inform decisions about instruction in the general education classroom so that all students are provided the opportunity to achieve well. Student progress monitoring, an essential feature of RTI, is a research-based practice that teachers can use to determine if they are teaching effectively and if their students are learning at an appropriate rate. When student progress monitoring is implemented effectively,

- Fewer referrals for testing are made, resulting in fewer inappropriate special education placements;
- Students who are identified for special services are done so more accurately based on their unique learning needs, rather than poor instruction;
- Struggling students are identified sooner and can be provided more effective instruction to meet their needs; and
- Teachers can plan their instruction so it meets the needs of each individual child and therefore the students' achievement increases.

Higher student achievement, lower rates of special education placement, and more accurate identification of students in need of more intensive instruction make progress monitoring a pivotal component of an effective, functional general education program

The rationale for developing the RTI model and the measures of its effectiveness are reflected in the New York State Performance Plan focus on quantifiable annual outcomes such as;

- **Addressing Student Achievement** Goal #3- Percentage of districts meeting the State’s annual yearly progress (AYP) objectives for progress for disability subgroups.

- **Addressing Disproportionality** Goal #10- Percentage of districts with disproportionate representations of racial and ethnic groups in specific disability categories that is the result of inappropriate identification.

- **Addressing Access to the General Education Curriculum** Goal #5-Percentage of children with IEPs ages 6-21 removed from regular classes less than 21 %, or over 60% of the day.

.

RTI uses evidence-based assessment practices to identify what students with poor academic skills **know** at initial screening, It then focuses on an **active learning process** to provide those critical skills and strategies that research has shown are necessary for mastery of basic reading, writing, spelling and math skills. It enables teachers to be proactive decision makers rather than self-limiting collaborators in a traditional system that is failing many students. Teaching, like learning, is an ongoing process of inquiry, in which the knowledge constructed about at-risk learners and their learning, as encountered in classroom situations, have the potential to continuously transform the teacher's way of understanding and acting in the classroom. Powerful interventions and progress monitoring can illuminate how responsive classroom instruction can be and suggest changes that help students achieve better.

RTI studies help examine how at-risk and not-at-risk students learn and the rate of mastery of basic strategies and skills needed to become proficient achievers. It can have impact only if it is based on the **best science available** and clearly demonstrates what works in helping students with poor academic performance achieve better in the general education classroom.

The promise of research on RTI, the evidence-based interventions and decisions about instruction provide a body of knowledge that can ultimately reshape how prevention and early intervention are designed and implemented in elementary schools. It can build capacity to offer all students a general education learning environment that enhances their opportunity to learn efficiently and effectively. Frequent assessment leads to skill improvement because matching students' needs with sound instruction makes instruction highly effective. Data-based decision making leads to both adjustments to instruction and more objective criteria for determining eligibility for special education services.

Many RTI studies use Curriculum-Based Measures (CBM) to examine essential academic skills. They help teachers monitor and compare rates of development under contrasting instructional interventions, identifying which instructional components result in optimal growth rates (see e.g., studies by Lynn and Doug Fuchs). CBM data can also help pinpoint the kinds of instructional programs and the academic skills specifically that students require to learn. Thus, teachers can change their instructional efforts to more efficiently implement only those evidence-based program components that actually result in good growth for students.

The practicality of the RTI model comes from an understanding that high quality instruction determines whether or not many at-risk students will be successful learners in school. Well designed, recent studies often show that students with experiential and instructional deficits can catch up to their not-at-risk peers and this group far outnumber the small percentage of students who are unable to profit from intensive and ongoing evidence-based interventions (see e.g., Velluntino et al., 2006).

A growing body of research on powerful interventions clearly locates the cause of reading difficulties (and consequential academic underperformance) in the child's educational experiences, and not in something deficient within the child. In other words, the child's capacity to learn to read is most often not the real issue (see Elrow et al., 2006). RTI helps replace teachers' assumptions about underachieving students with accurate, evidence-based information Studies, e.g., have addressed efforts to maximize students' attention to task, academic engaged time, academic learning time, and ultimately, their specific and generalizeable learning on critical tasks. It also can create a generalized response to underachievement reflecting a school's greater responsibility for learning. RTI research should help transfer the focus of scrutiny and burden of adjustment from students to the materials and methods of instruction encountered in the general education classroom.

.

RTI research attempts to identify the most powerful diagnostic interventions that demonstrate what can and needs to done in the general education classroom. It establishes what specially designed instruction works for individuals by engineering a match between students' current skills and strategies and research-based interventions. It provides the evidence and information that can influence teachers' beliefs and behaviors concerning teaching, learning, and students with poor academic performance. RTI suggests how to accurately identify children who, without specially designed instruction (Tier-2 small group instruction), would develop reading disabilities and require special education services.

RTI research is important because it addresses quality teaching, the context for learning, student behaviors and service delivery outcomes, helping untangle and unpack their complex connections in general education.

II. Second Generation RTI research

A. Making Policy Work

This synthesis of research considers studies conducted after the IDEA legislation of 2004, as a decided change in the focus of research has occurred. RTI is not a new concept. It has built on an extensive body of research findings about reading (see National Reading Panel, 2000), the limitations of the refer-test-place process over a two decode timeframe, and an urgency to reform general education (Reading First and NCLB, 2001). In effect, science is trying to catch up to social policy. The OSEP goals to reduce the growing numbers, and minority overrepresentation, in special education, to engage more students with disabilities in the general education curriculum, to make that curriculum more effective with all students so fewer will be struggling to become functionally literate, are pushing research to be more practical and offer solutions to seemingly intractable problem in general and special education.

New US government requirements state that federally funded grants and school programs must prove that they are based on scientifically proved improvements in teaching and learning. All new grants must show they are based on scientifically sound research to be funded, and budgets to schools must likewise show that they are based on scientifically sound practices.

The studies reviewed in this synthesis go beyond the pioneering work that persuasively indicated, e.g., that formative assessment and progress monitoring are superior to one shot summative testing, that authentic tests of curriculum used in classrooms are more helpful that tests developed to provide nationally standardized grade scores, that

differentiated instruction is more responsive to students needs than undifferentiated whole class instruction, and that teaching reading skills systematically and explicitly with modeling, immediate feedback, reward and extensive practice, are learned better than any other way. The pioneering research is comparable to Levin and O'Donnell's (1999) stages 1 and 2 of their four stage model, whereas second generation RTI research is analogous to stages 3 and 4.

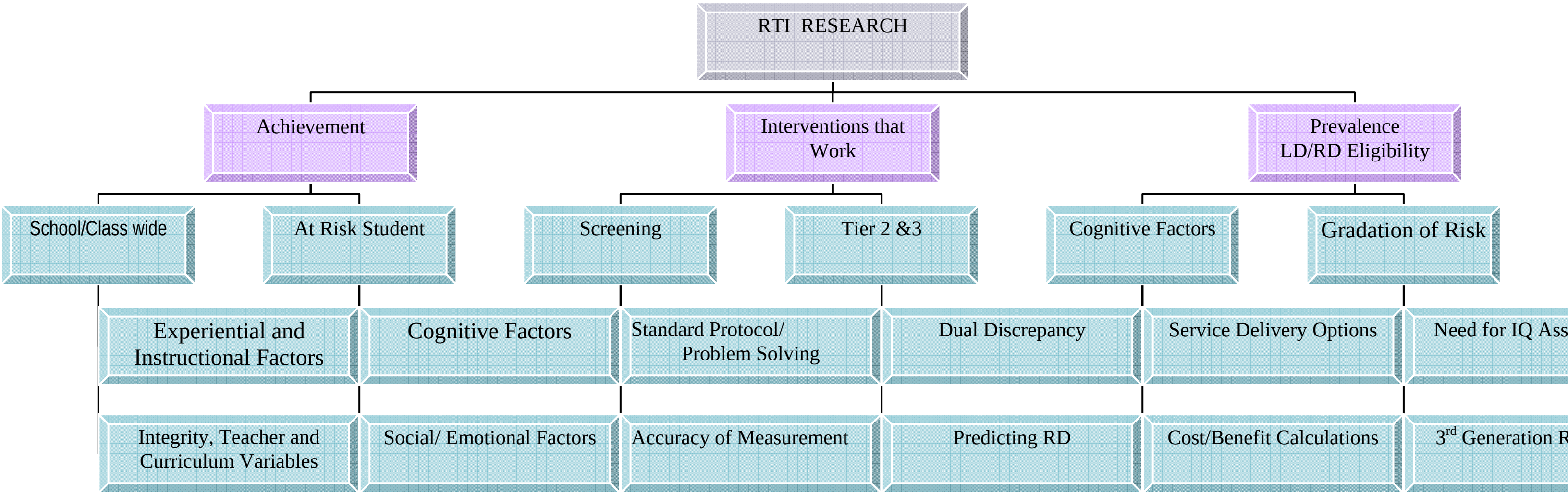
Stage 1- Preliminary ideas, review of research, hypotheses and pilot work pertinent to the development of interventions.

Stage 2- Test the effectiveness of their intervention via controlled laboratory experiments or through classroom-based demonstration and design experiments (preliminary but critical to inform classroom practice, see stage 4).

Stage 3- Randomized classroom trials; realistic and carefully controlled conditions are administered long enough for the intervention to take place; both desired and unwanted side effects can be assessed; and threats to internal validity is controlled by experimental procedures.

Stage 4- Informed classroom practice. The application of research-validated interventions by classroom teachers meet three conditions: are knowledgeable of the intervention; decide that their students and classroom contextual variables are an appropriate match to the intervention and validating data; and must implement and evaluate the effects of the intervention with their own students (fidelity of implementation, screening and ongoing progress monitoring).

The current generation of research builds on a well established, valid and reliable intervention strategies, powerful screening and outcome assessment measures, and sensitive predictors of at-risk and learning and social behaviors, to answer questions about how to **operationalize** the extant science-based knowledge in schools.


RTI RESEARCH
Achievement
Interventions that Work
Prevalence LD/RD Eligibility
School/Class wide
At Risk Student
Screening
Tier 2 &3
Cognitive Factors
Gradation of Risk
Experiential and Instructional Factors
Cognitive Factors
Standard Protocol/ Problem Solving
Dual Discrepancy
Service Delivery Options
Need for IQ Ass
Integrity, Teacher and Curriculum Variables
Social/ Emotional Factors
Accuracy of Measurement
Predicting RD
Cost/Benefit Calculations
3rd Generation R

There are three main directions of second generation research (a) improving general education instruction, (b) understanding how to intervene to address at-risk learning and social behavior, that left unaddressed, leads to school failure, and (c) decision-making about LD eligibility.

Current research deals with a wide range of issues reflected in the broad categories in the chart above, including: scaling up single subject research findings to school-wide, district and state levels; cost/benefit realities investigating which materials, strategies, assessments are most efficient and effective when resources are limited; public health models that provide the greatest impact for prevention and early intervention efforts; and how to identify students at-risk students early enough to put practices in place that will make a difference in their lives.

B. Making Research Credible

School practitioners are not researchers. Stake holders need to be persuaded that this time, this reform movement, is worth buying into. The urgency to change practice is based on science that is as close to specifying cause and effect relationships as is possible in the context of schools. Good science can explain how and why students at-risk learn the way they do and what schools can do to enhance that learning. Current rigorous research guidelines require RTI researchers to be able to: compare the experimental treatment with an appropriate alternative or nonintervention condition; demonstrate that expected outcomes are produced under varied conditions in replication studies; show direct relationships or connections exist directly between the treatment and the specified results; and eliminate any reasonable competing explanation for the obtained outcomes (usually through randomization and methodological safeguards).

III. Study Outcomes

1. Effect Size and p Values
Ten out of thirty-one studies report data in terms of effect size. The strongest findings were reported in three studies involving early grades phonemic awareness components that predicted later reading deficiency: ES= 1.40 for letter-sound fluency (Speece); ES= 1.30 for phoneme segmentation (Velluntino et al.); and ES= 2.87 for sound blending (Roth et al.). These findings are consistent with Elbrow and Peterson's moderate effect size results (in the range from .40 to .69) for phoneme awareness training that was associated with positive long-term effects in reading in grades 2, 3, and 7.

Findings in the other twenty-one studies were reported using only p values of .05 or greater, probably in part, because ES is a more stringent test of significance than are probability values. Furthermore, few peer-reviewed journals require ES data reporting, and as noted in Appendix A, only one journal out of eleven sampled recently (School Psychology Quarterly) published an issue where most articles provided effect sizes.

2. LD/RD Eligibility and Risk Factors
Several second generation studies bring new concepts and data to address questions of who is learning disabled, and can we reduce the incidence of LD by making better service delivery choices? Velluntino et al.'s description of **Gradation of Risk** is an important concept in LD identification. It is a continuum determining how easy it will be for a student to acquire functional literacy based on measures of instructional, experiential and biological factors. These researchers showed that the vast majority of poor reading kindergarten students will learn well and catch up to average reading peers with evidence-based interventions teaching early literacy skills. However, difficult to remediate kindergarteners may need to be considered for special education services. This conclusion is based on following students through two years of intensive intervention work where a combination of low cognitive skills related to reading, and poor experiences make it unlikely that the general education classroom can offer them intensive enough services to become functionally literate.

McDermott et al.'s regression analysis study identifies males, minority members, students with poor verbal and nonverbal cognitive skills, students with behavior problems, and those with processing speed problems as at high risk to be identified as LD. The research sensitizes us to objectively use interventions for students most likely to be referred by classroom teachers. Lipka examined and found evidence for a "fourth grade slump" among an unexpectedly large number of poor reading fourth graders. These students tested as average readers in earlier grades. Lipka suggested that basic skills not learned to automaticity will create difficulties when phonologically and morphologically complex words overwhelm marginal basic reading skills. It is important to consider that some students will need ongoing progress monitoring and formative assessment in upper grades to assess their rate of progress even if previously not identified as at-risk

Klinger et al. and Hintze et al. investigated the impact of procedural integrity when teachers use interventions and caution that poor integrity weakens the ability of poor readers to profit from interventions. Teachers point to time management issues and competing priorities for lack of rigorous implementation of interventions. A problem inherent to standard protocol research is whether well designed interventions are practical in the classroom as they are developed under laboratory conditions.

Compton et al.'s 2006 and Velluntino et al.'s studies use a standard CBM battery to identify at-risk students in kindergarten and first grade. The battery includes measures of phonemic awareness, oral vocabulary, rapid naming, and initial word identification tasks. Each of the cognitive measures has essential but not sufficient predictive capabilities in and of themselves. The battery is a product of considerable early RTI research establishing powerful predictors of later reading success.

Second generation studies thus build upon the successive strides of prior small group intervention studies and can now reliably use interventions and test batteries to maximize RTI's ability to identify students who, without assistance, will become reading disabled. In the same manner, second generation studies with larger samples, effect size data permitting cross study comparisons, randomization, broader learning variables scrutinized, longitudinal retrospective designs, and in-classroom variables considered, will further refine and enhance RTI research.

What is an 'Effect Size'?

A brief introduction

Robert Coe, CEM Centre, Durham University

March 2000

'Effect Size' is simply a way of quantifying the difference between two groups. For example, if one group has had an 'experimental' treatment and the other has not (the 'control'), then the Effect Size is a measure of the effectiveness of the treatment.

Effect Size uses the idea of 'standard deviation' to contextualize the difference between the two groups. Standard deviation is a measure of how spread out a set of values are. Various formulae for calculating it can be found in any statistics text book, or if data are entered into a spreadsheet such as Excel, a built-in formula can be used.

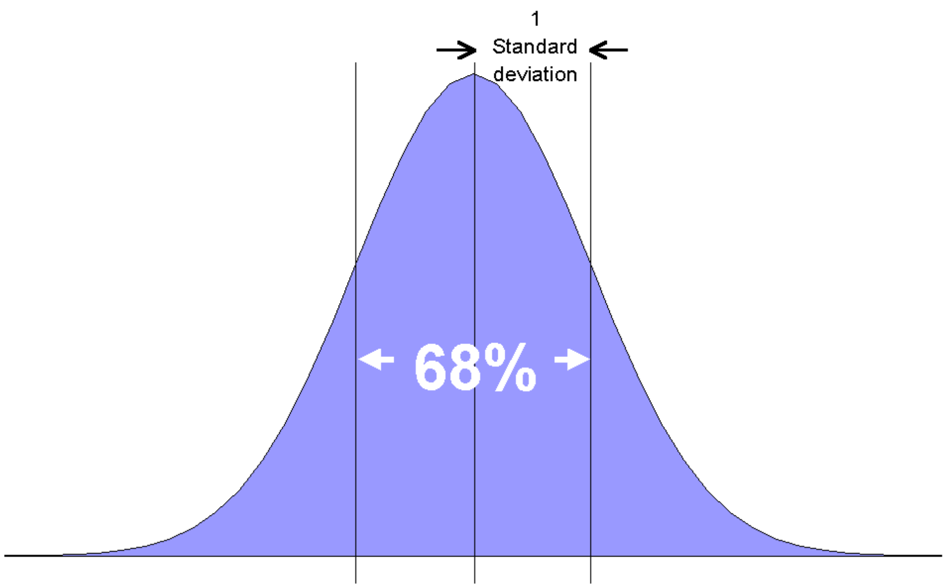


Figure 1

Alternatively, standard deviation can be interpreted graphically. Many datasets have a distribution similar to that shown in Figure 1. For these, the standard deviation is the distance you have to go either side of the mean (average) in order to include 68% of the population. If you go twice this distance (two standard deviations), then you can expect to include 95% of the population.

The Effect Size is just the difference between the mean values of the two groups, divided by the standard deviation (Equation 1).

$$\text{Effect size} = \frac{\text{Mean of experimental group} - \text{Mean of control group}}{\text{Standard deviation}}$$

Equation 1

An example:

Consider an experiment conducted by Val Dowson to investigate time of day effects on learning: do children learn better in the morning or afternoon? A group of 38 children were included in the experiment. Half were randomly allocated to listen to a story and answer questions about it at 9am, the other half to hear exactly the same story (on tape) and answer the same questions at 3pm. Their comprehension was measured by the number of questions answered correctly out of 20.

The average score was 15.2 for the morning group, 17.9 for the afternoon group: a difference of 2.7. But how big a difference is this? If the outcome were measured on a familiar scale, such as GCSE grades, interpreting the difference would not be a problem. If the average difference were, say, half a grade, most people would have a fair idea of the educational significance of the effect of reading a story at different times of day. However, in many experiments there is no familiar scale available on which to record the outcomes. The experimenter often has to invent a scale or to use (or adapt) an already existing one – but generally not one whose interpretation will be familiar to most people.

Using Effect Size helps to overcome this difficulty, since if we know the spread of scores (ie the standard deviation), it will help us to put the difference into context. In Dowson's time-of-day effects experiment, the standard deviation (SD) = 3.3, so the Effect Size was (17.9 – 15.2)/3.3 = 0.8.

Interpreting Effect Sizes

Provided our data have the kind of distribution shown in Figure 1 (a 'Normal' distribution), we can readily interpret Effect Sizes in terms of the amount of overlap between the two groups.

For example, an effect size of 0.8 means that the score of the average person in the experimental group exceeds the scores of 79% of the control group. If the two groups had been classes of 25, the average person in the 'afternoon' group (ie the one who would have been ranked 13th in the group) would have scored about the same as the 6th highest person in the 'morning' group. Visualizing these two individuals can give quite a graphic interpretation of the difference between the two effects.

Table 1 shows conversions of effect sizes to percentiles (column 2) and the equivalent change in rank for the average person in a group of 29 (column 3). Notice that an effect-size of 1.6 would raise the average person to be level with the top ranked individual in the control group, so effect sizes larger than this are illustrated in terms of the top person in a larger group. For example, an effect size of 3.0 would bring the average person in a group of 740 level with the previously top person in the group.

Table 1: Interpretations of effect sizes

Effect Size	Percentage of control group who would be below average person in experimental group	Rank of person in a control group of 29 who would be equivalent to the average person in experimental group
0.0	50%	15
0.1	54%	13
0.2	58%	12
0.3	62%	11
0.4	66%	10
0.5	69%	9
0.6	73%	8
0.7	76%	7
0.8	79%	6
0.9	82%	5
1.0	84%	5
1.2	88%	3
1.4	92%	2
1.6	95%	2
1.8	96%	1
2.0	98%	1 (or 1^{st} out of 44)
2.5	99%	1 (or 1^{st} out of 160)
3.0	99.9%	1 (or 1^{st} out of 740)

Another way to interpret effect sizes is to compare them to the effect sizes of differences that are familiar. For example, an effect size of 0.2 corresponds to the difference between the heights of 15 year old and 16 year old girls in the US. A 0.5 effect size corresponds to the difference between the heights of 14 year old and 18 year old girls. An effect size of 0.8 equates to the difference between the heights of 13 year old and 18 year old girls.