Response to Intervention: Considerations for Research and Systems Change

Joseph F. Kovaleski
Indiana University of Pennsylvania

This invited commentary on the series of articles regarding the state of the extant research and practice concerning the implementation of response to intervention (RTI) will focus on the following issues: (a) what constitutes a comprehensive service delivery system that uses RTI as its driving principle, (b) the dependent measures by which the effects of RTI should be evaluated, (c) how treatment fidelity can be meaningfully facilitated in real school environments, (d) what can be done with students who are not responsive to even robust interventions, and (e) which students should be identified as having specific learning disabilities. Throughout the discussion of these issues, implications for the practice of using RTI in the schools and research directions will be presented.

The Multitier Service Delivery Model

A worthwhile place to start is to attempt to offer some clarity about what is being proposed as the multitier service delivery model that uses RTI as its recurrent scheme to design interventions, monitor progress, and make both small-scale and large-scale decisions about students. In many descriptions of RTI, both here (Glover & DiPerna, 2007) and elsewhere, it appears to be assumed that the projects that began in the 1990s, including Iowa's problem-solving model (Tilly, 2003), Pennsylvania's instructional support teams (Kovaleski & Glew, 2006), Rosenfield and Gravois's (1996) instructional consultation teams, and others are being presented as intact RTI models. However, although these precursors used problem solving and progress monitoring as core components of their models, none of them embodied all of the components of current conceptualizations of a multitier service delivery system (e.g., Batsche et al., 2006). These pioneering approaches to intervention support models, which originated before the advent of the National Reading Panel (2000), standard protocol interventions (Wanzek & Vaughn, 2007), school-wide positive behavior supports (Sugai & Horner, 2002), the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2005), and other research-based initiatives, generally assumed (or did not address) the effectiveness of the general education program and endeavored to customize interventions for individual
students who were not successful in it. (It is acknowledged that some of the beginning tiers or phases of these models involved consultation with general education teachers, but not for the purpose of making fundamental changes to the core program.) They were, in effect, single-tier approaches, and in today’s terms might be understood somewhere in the middle of most contemporary multitier RTI models. Simply stated, RTI cannot be equated with the problem-solving process or with problem-solving teams.

Nonetheless, as late as 2003, Fuchs, Mock, Morgan, and Young were posing the problem-solving process and standard protocol interventions as alternative approaches to RTI. In contrast, current depictions of a multitier service delivery model (Batsche et al., 2006) incorporate both of these approaches as fundamental components. Furthermore, in an effort to “harvest” the research efforts of the last 10 years, other features such as instituting a scientifically based core curriculum; universal screening with DIBELS, AIMSweb (Shinn & Germann, 2006), and other instruments; data-analysis teaming; differentiated instruction in general education; and school-wide positive behavior supports have been proposed as fundamental components of a comprehensive model (Kovaleski, 2007). The most important distinction between the precursor and current models is that the effectiveness of the core academic and behavior programs for all students are not taken for granted. Rather, in the face of large numbers of students who continue to fall below proficiency levels, the current models directly address the source of the problem: the inability of many schools to bring large percentages of their students (80% according to Batsche et al., 2006) to proficiency in basic skills. Per an old metaphor, we have moved our efforts from downstream to upstream.

This new focus on the core program of curriculum and instruction has important implications for not only how RTI is conceptualized, but also how we evaluate the extent to which it is implemented (treatment fidelity) and how effective it is in terms of outcomes for students, which will be addressed next.

The RTI Research Agenda

At this stage of development in the use of RTI in the schools, it is appropriate for research to focus on the individual components of the multitier service delivery process and on RTI as a measurement tool itself (Glover & DiPerna, 2007). Some of this analysis is underway (although certainly not completed), as indicated by the articles in this series on universal screening (Jenkins, Hudson, & Johnson, 2007), progress monitoring (Shinn, 2007), extensive interventions (Wanzek & Vaughn, 2007), skill acquisition and generalization (Daly, Martens, Barnett, Witt, & Olson, 2007), and professional development (Kratochwill, Volpiansky, Clements, & Ball, 2007), as well as in other publications (cf. Barnett et al., 2006). Other proposed features of the multitier model (e.g., using data-analysis teams to assist with implementation of the core curriculum) are promising approaches (Kovaleski & Pedersen, in press), but do not yet have research support as individual components. However, two more general lines of research have not thus far been undertaken: (a) research on overall effects of various core curriculum “packages,” and (b) research on the effects of the comprehensive model when implemented in its entirety.

It is perhaps surprising that schools’ curricular choices, a central aspect of schooling that entails the “what” that students are expected to learn, have not been systematically studied. By definition, a curriculum consists of a school’s scope and sequence of knowledge and skills to be learned, and is operationalized by the materials and methods used to deliver that content. Although it would be erroneous to equate the materials that schools purchase and use with their curriculum, it is well understood that the materials teachers use strongly influence what happens on a minute-to-minute basis on any given school day. However, obvious questions such as whether a given set of curriculum materials (e.g., Houghton-Mifflin Reading, Wilson Reading) produces acceptable levels of overall student proficiency (e.g., 80%) across the many schools that use the series have rarely been
empirically addressed (with the possible exception of studies of Direct Instruction). There have been helpful attempts by the Florida Center for Reading Research (http://www.fcrr.org) and the University of Oregon’s Institute for the Development of Educational Achievement (http://reading.uoregon.edu/) to analyze and rate these curriculum products in terms of their coverage of the “big ideas” of reading, but whether the faithful use of these materials results in positive student outcomes is still largely unknown.

What is needed is a line of research that parallels the one described by Wanzek and Vaughn (2007), in which standard protocol interventions have been empirically evaluated. In essence, we now have a good beginning on the evaluation of Tier 2 and 3 interventions, but no idea about what it will take to get the core curriculum to work at Tier 1. A complicating issue with this potential line of research is that many schools use multiple materials as their core program. For example, one low-to-middle income elementary school in western Pennsylvania supplements Everyday Math with Mastering Math Facts, and regularly realizes proficiency rates above 90% in mathematics on the state tests. Although random assignment of curricular materials to school districts will not be possible, it would perhaps be fruitful to study schools that use various combinations of both general and supplemental materials as regular parts of their core program. For example, one low-to-middle income elementary school in western Pennsylvania supplements Everyday Math with Mastering Math Facts, and regularly realizes proficiency rates above 90% in mathematics on the state tests. Although random assignment of curricular materials to school districts will not be possible, it would perhaps be fruitful to study schools that use various combinations of both general and supplemental materials as regular parts of their core program (e.g., Ysseldyke, Spicuzza, Kosciolek, & Boys, 2003). The extent to which these materials are used with fidelity will need to be factored in to this analysis. In summary, to evaluate the effects of RTI and the multitier model, it is first important to understand the effect of the broad base of the pyramid at Tier 1.

The second critical line of research into the effects of RTI is the extent to which the entire comprehensive model produces positive outcomes when all individual components are implemented with effectiveness and when the process is operated with an appropriate level of efficiency. This analysis presupposes verification that all components have been implemented with fidelity, which will be addressed in the next section. Assuming good treatment integrity, there is the question as to the appropriate dependent measure to appraise the effectiveness of the overall model. Vaughn and Fuchs (2006) have properly noted that a decrease in the number of students in special education is not an appropriate outcome for evaluating the effectiveness of RTI. Similarly, Kovaleski and Glew (2006) noted that although stemming the rapid rise of students identified with specific learning disabilities (SLD) was frequently used as a rationale for instituting various “regular education initiatives” in the 1990s, the passage of the No Child Left Behind (NCLB; 2001) legislation created a new societal goal—increasing the percentage of students who pass state proficiency tests, with a terminal goal of 100% proficiency for all students by 2013–2014. Although descriptions and discussions of RTI currently populate the special education and school psychology literature, the correct placement of the multitier model is within general education. From a research perspective, then, the needed line of study is the extent to which the multitier model results in increasing percentages of students passing state tests, with improvements in indicators of effective teaching analyzed as a mediating variable. The success of the multitier model will be judged not only by the extent to which the core curriculum (in Tier 1) brings increasing numbers of students to proficiency, but also the extent to which the added procedures (e.g., data-analysis and problem-solving teaming) and extensive supplemental programs (i.e., standard protocol interventions) increase those percentages toward the overall goals set by NCLB. In essence, the research on RTI needs to include continued investigations into the efficacy of various individual components as well as into the overall effectiveness of the program when all the components are brought together in a coherent service delivery system.

Maximizing Treatment Fidelity

Ensuring sufficient treatment fidelity as a sine qua non of RTI has been extensively described in this series by a number of the
authors. The prospect of using RTI as part of a comprehensive evaluation for the diagnosis of SLD presupposes that the research-based interventions have actually been carried out with a high level of implementation. As indicated by the authors of this series, achieving these levels will require changes in preservice and in-service training, and a clear understanding of who should monitor and facilitate treatment fidelity in the schools.

The arguments made by Kratochwill et al. (2007) regarding the lack of transportability of research-based interventions into educational environments are reminiscent of seminal studies by Flugum and Reschly (1994) and Telzrow, McNamara, and Hollinger (2000) regarding the lack of implementation of strategies recommended by problem-solving teams in actual classroom use. Wanzek and Vaughn (2007) further note that there are no extant studies of the effectiveness of interventions generated through a problem-solving process. Further darkening the current educational landscape is a disturbing study by Walsh, Glaser, and Wilcox (2006), who documented the extensive variability among colleges of education in training teachers in the big ideas of reading. In a nationwide sample, some university programs show no (0%) coverage of these concepts and skills, whereas others have high rates of coverage (to 100%). Finally, Kratochwill et al. (2007) also note the "antisience" stance of many educators, which further impedes efforts to implement research-based practices in the schools.

The resolution of these difficulties, which affect not only the implementation of RTI but also the ability of schools to meet the challenges of NCLB, may lie in three areas: intense and ongoing training, collaborative support structures, and administrative follow-through. First, as indicated by Glover and DiPerna (2007), training efforts need to be frequent and of sufficient intensity to allow teachers to build the requisite skills. The training also needs to be durable so that initial training is followed up with opportunities for guided practice by expert teachers and/or consultants as well as opportunities for staff to discuss issues with implementation. Glover and DiPerna's description of how these functions are embodied by positive behavior support leadership teams is a good example of this type of ongoing training.

Kratochwill et al. (2007) describe structures to support teachers in their implementation of new instructional tactics (e.g., teacher networks and study groups). In the same vein, Kovaleski and Pedersen (in press) have detailed a step-by-step procedure for instituting and operating data analysis teams in which teachers review results of universal screening to plan instructional practices with the goal of bringing an increased number of students to proficiency in basic skills. Their preliminary results indicated that schools using these teams become increasingly successful in meeting their goals. Most interesting has been the report of these teachers on the acceptance of the suggested strategies. In contrast to the well-established principle that teacher acceptability is a preliminary step to implementation (Witt & Elliott, 1985), it appears that teachers may become more accepting when interventions actually work with their students. This preliminary observation mirrors Skrtic's (1991) assertion that paradigm change requires the observation of anomalous information. In this case, when teachers are surprised that even students who have deficient skills, or who come from low-income environments, acquire the needed skills when robust interventions are used, they become more accepting of the interventions. Further research on this notion is certainly warranted.

A related question is who is really responsible for treatment fidelity. Interestingly, the concept of treatment fidelity is pervasive in the literature of school psychology and special education, but appears not to be a large concern in publications related to school administration. However, it is fair to propose that the school principal is the person responsible for ensuring that school-wide programs, including the core curriculum and supplemental interventions, are delivered with fidelity. The existing literature in the area of consultation is fraught with the image of consultants (e.g., school psychologists) working to convince "consultees" to implement needed interven-
tions, which implies a clear sense of optional-
ity in following through on the recommenda-
tions. However, because NCLB requires that
schools use research-based practices, perhaps
it is time to move from recommending good
ideas to expecting a certain level of profes-
sional performance. To make this shift will
require a cross-fertilization of research from
one domain of education to another. For ex-
ample, school principals should become famil-
iliar with the work of Noell and his colleagues
(e.g., Noell & Gansle, 2006), which has dem-
onstrated that information feedback facilitates
the implementation of treatment plans. Such
work can inform principals as they review
lesson plans (the practical daily equivalent of
treatment plans), observe teachers delivering
lessons, and provide feedback, all of which are
traditional, if not well-directed, administrative
functions. Glover and DiPerna (2007) hint at
this level of change when they describe insti-
tutional changes to policies and practices. In
summary, efforts to ensure treatment fidelity
should be a combination of intensive and du-
rable preservice and in-service training, col-
laborative structures that support teachers in
learning and implementing new programs and
strategies, and clear administrative expecta-
tion setting and monitoring.

Addressing the Needs of Nonresponsive
Students

Proponents of RTI have been heartened
by the generally positive results from studies
in which extensive interventions (i.e., standard
protocols) have been shown to increase stu-
dents’ skill on essential basic skills (Wanzek
& Vaughn, 2007). The large effects noted
from kindergarten to third grade hold promise
that these gains can be maintained for these
students as they progress through their educa-
tional careers. It is also apparent from these
studies that there are certain students who do
not respond well to intervention packages that
work for the majority of students. How to
address the needs of these students is an active
debate within the school psychology literature.
What seems to be universally acknowledged is
that it is probably wise to start with evidence-
based practices that have been shown to work
with large percentages of students who have
similar needs (e.g., deficiencies in phonemic
awareness). As Reyna (2004) has indicated,
“(d)espite the intuitive appeal of the familiar
slogan, ‘one size does not fit all,’ some edu-
cational practices are broadly effective; they
can be generalized widely across contexts and
populations” (p. 56). Both sides of the debate
also appear to agree that when students fail to
respond to typically effective interventions,
attempts should be made to customize an in-
tervention for the individual student. As indi-
cated earlier, problem-solving teams have had
this individualization as a core practice for
many years. However, as Wanzek and Vaughn
(2007) indicated, it is not clear what types of
interventions these teams select or whether
they are effective. Both sides would also agree
that the design of an intervention at this point
should follow a more extensive assessment,
and it is here that the debate is joined. The two
positions can be described as the diagnostic-
prescriptive approach and the ecological-skills
approach.

The diagnostic-prescriptive approach
has been an historic one in special education
and school psychology (Reschly and Ys-
seldyke, 2002) and is embodied in current
form by those who seek explanations for
nonresponsiveness in deficiencies of cogni-
tive processes (Hale, Kaufman, Naglieri, &
Kavale, 2006). Interventions in this approach
would address or accommodate for these pro-
cesses. The ecological-skills perspective, as
comprehensively described by Daly et al.
(2007) and other authors (e.g., Howell & No-
let, 2000) takes a task-analytic (behavioral)
approach to understanding the discrete skills
needed for the student to respond correctly and
analyzes the student’s skill deficiencies in
making these responses. From a research per-
spective, the empirical test for both ap-
proaches is whether the described additional
assessments lead to interventions that (a) are
different than what would have been identified
based on the data already collected during the
early tiers, and (b) lead to better outcomes for
the nonresponsive students than would have
been realized merely by increasing the fre-
quency or intensity of the nonproductive strategies. Because NCLB requires that schools track the proficiency of even the most intractable learners (i.e., those with disabilities), the outcomes of these future research efforts with nonresponding students will be eagerly awaited by school practitioners.

Measurement Issues

This final set of comments relates to a number of measurement issues raised in this series, particularly to selecting universal screening instruments (Jenkins et al., 2007) and using RTI as part of the eligibility decision-making process (Shinn, 2007). First, it has been encouraging to watch the proliferation of DIBELS, AIMSweb, and other instruments as schools begin to take seriously the achievement of all students. One reason for the popularity of these instruments is that they provide school administrators with a glimpse of how students will perform on the less frequently administered (annual) state tests. Consequently, as formative indicators, the gold standard for these measures is their ability to predict performance on the state tests. To this point, research seems to indicate that measures of oral reading fluency are good predictors of state test performance (Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). However, it is less well known whether oral reading fluency holds up as a predictor for older students. One would speculate that measures appraising a student’s comprehension of text more directly would be better formative measures in later grades, as the state tests increasingly tap comprehension at these levels. One such formative measure is 4Sight (Slavin & Madden, 2006), which has displayed high correlations with state test results in studies conducted to date (Success for All Foundation, 2007).

School psychologists are most interested in how data collected regarding a student’s RTI might inform the eligibility decision-making process. Shinn (2007) notes that students currently in special education are those with the most deficient skills (rather than those with ability–achievement discrepancies per se), and proposes that the use of RTI will therefore not result in a different type of student being identified as SLD. However, there will likely be a great deal of discomfort among practitioners regarding his proposal to use relative achievement discrepancies (using local norms) as the basis for determination of eligibility. This discomfort will have more to do with people’s sense of fairness about who receives special education across districts than with concerns about the purity of the construct of SLD. Further, in conceptualizing the cut score against which a given student’s performance should be compared, it can be argued that the Individuals With Disabilities Education Act (2006) gives clear direction:

The child does not achieve adequately for the child’s age or to meet state-approved grade-level standards in one or more of the following areas, when provided with learning experiences and instruction appropriate for the child’s age or State-approved grade-level standards. . . . (§300.309[a][1])

This provision indicates that the benchmark against which a student’s deficiency is determined should be the student’s performance in relation to the state (not local) standards. Operationally, this metric would be the state test, and the relevant cut score would be the lowest score in the proficient range. How far below that level a student needs to be for special education eligibility has not been defined, and will likely create a good deal of debate at the state and local level, regardless of whether local education agencies (LEAs) choose the RTI option. However, if LEAs follow this requirement, it does appear that who qualifies as SLD may change, especially in high-performing districts that currently have substantial numbers of identified students who may be discrepant from their local peers but not deficient in relation to state standards. This regulation implies that students who do not perform substantially below what would be expected in relation to state standards do not qualify for special education services (regardless of their assessed IQs). It is possible that this provision will do more to change who is in special education than the adoption of RTI.

643
LEAs that adopt RTI will be faced with other measurement conundrums. If the robust interventions suggested by RTI are implemented with fidelity, many districts will be faced with the prospect of students who are very deficient, yet capable of making meaningful progress. However, for students who are very deficient, even accelerated rates of improvement will not immediately bring them to an acceptable level of proficiency. Consequently, assessment teams will need to monitor students’ progress with an eye toward the trajectory of that progress over years rather than the biannual or annual perspective used with most curriculum-based measurement applications. Deno and Mirkin (1977) introduced a multiyear progress monitoring format that provides for such long-term performance tracking and that may have renewed utility 30 years after its publication. In essence, the question will be how durable and persistent LEAs will be in implementing evidence-based practices that are working with individual students, in some cases needing to maintain appropriate levels of frequency and intensity for many years. This prospect presents particular challenges for secondary schools that have frequently not seen their mission as addressing the needs of lagging students who are not eligible for special education.

A related issue here again is the need for training in curriculum-based measurement and other assessment methods with those teachers and specialists who will be implementing the interventions. It will not be sufficient for school psychologists and special education teachers to have these skills. Title 1 and other remedial specialists will need to embrace these methods. Regrettably, although curriculum-based measurement is pervasive in the school psychology and special education literature, it has not yet been widely disseminated in journals read by these teachers. Efforts to break out of practitioners’ educational “silos” will be needed to accomplish these tasks.

Summary

Schools that desire to implement RTI and scholars who endeavor to study its effects are in a period of great opportunity. Despite the many research questions that need to be addressed, there is a sufficient evidence base regarding many of the components of a comprehensive, multitier system to warrant adoption of the procedures at a local level. It is argued that, although these components should be implemented in a planful and sequential manner, LEAs should not imagine that they are sufficiently implementing RTI until all the components are in place and operating in an efficient manner. Research into how these individual components (especially the effects of various core curricular programs) as well as the model as a whole affect improved teaching and, most important, the improvement of students’ basic skills, will need to be addressed in the coming years. Meanwhile, both practitioners and researchers will strive toward a better understanding about how treatment fidelity can be realistically accomplished, how nonresponsive students can be effectively taught, and which students among those who fail to respond adequately qualify for special education.

References


Joseph F. Kovaleski, DEd, NCSP, is Professor of Educational and School Psychology and Director of Graduate Programs in School Psychology at the Indiana University of Pennsylvania. His research interests include system-wide efforts for school restructuring and the implementation of response-to-intervention programs in the context of whole-school reform.